“We learn from failure, not from success!”

-Bram Stoker, *Dracula*
Augmented Biofeedback for Partial Weight-Bearing Learning

by

Ian Smith

BCS, BSc, University of New Brunswick, 2017

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

Master of Computer Science

in the Graduate Academic Unit of CS

Supervisor: Scott Bateman, PhD, Faculty of Computer Science
Erik Scheme, PhD, Department of Electrical Engineering

Examining Board: Michael Fleming, PhD, Faculty of Computer Science
Suprio Ray, PhD, Faculty of Computer Science
Kevin Englehart, PhD, Department of Electrical Engineering

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

August, 2019

©Ian Smith, 2019
Abstract

Assistive devices, including canes and crutches, are used in partial weight-bearing (PWB)—offloading weight from limbs weakened by disease or injury to promote recovery and prevent reinjury. While it is important to accurately offload weight to target loads prescribed by healthcare providers, current training methods result in poor compliance. It is currently unknown how to most effectively provide feedback during training to allow users accurate execution of the skill later on. In this work, three studies were conducted to investigate the effects of feedback modality, delay, and resolution on both regulation and learning of PWB while stationary and during gait. Results indicate that concurrent feedback is best suited for continuous skill regulation whereas retrospective feedback is preferable for training PWB, and that task-specific training is critical for compliance. This work presents design guidelines for improved clinical PWB training methods and highlights the importance of researching retrospective motor learning methods.
Acknowledgements

The proverb “it takes a village to raise a child” certainly applies to this thesis. I would like to thank my friends and family for the support they’ve offered throughout this process. Whether it be putting up with my cane puns or a night of unwinding after a stressful deadline, you were and continue to be there for me.

The Institute of Biomedical Engineering has been incredibly welcoming of me despite my lack of Iron Ring, a gesture that I sincerely appreciate. Furthermore, I am thankful for Dr. Satinder Gill’s tireless efforts on the Smart Cane project, especially his willingness to partake in Saturday morning debugging sessions.

Thank you to my fellow basement-dwellers in the UNB HCI lab for fostering a phenomenal culture of collaboration and cooperation, along with occasional shenanigans to keep academia interesting.

I would be remiss in not thanking NBIF and NSERC for making this research possible through their generous funding.

Finally, I’d like to thank my supervisors Dr. Scott Bateman and Dr. Erik Scheme. I walked into this project without a clue what I was doing, yet you somehow guided me to find purpose in my academic pursuits.
# Table of Contents

Abstract .............................................................................................................................. ii
Acknowledgements ........................................................................................................... iii
Table of Contents ............................................................................................................. iv
List of Figures ..................................................................................................................... x
1 Introduction ..................................................................................................................... 1
2 Background ....................................................................................................................... 4
   2.1 Mechanics of Gait ........................................................................................................ 4
   2.2 Gait monitoring methods ......................................................................................... 7
   2.3 Partial Weight-Bearing ............................................................................................ 9
   2.4 PWB Training ........................................................................................................... 11
   2.5 Internal Models ........................................................................................................ 12
   2.6 Guidance .................................................................................................................. 15
   2.7 Biofeedback for PWB ............................................................................................. 16
3 Stationary Loading Study ............................................................................................... 21
   3.1 Motivation ................................................................................................................ 21
   3.2 Experiment-Specific Background .......................................................................... 22
      3.2.1 Feedforward Control in Partial Weight-Bearing .............................................. 22
      3.2.2 Feedback for PWB Training ............................................................................. 24
      3.2.3 Feedback Design for Training Motor Control Tasks ....................................... 26
   3.3 Feedback Design ....................................................................................................... 27
      3.3.1 Feedback Schemes ......................................................................................... 29
3.4 Experiment ................................................................................................................................... 30

3.4.1 Participants................................................................................................................................ 31

3.4.2 Experimental Task ............................................................................................................................. 32

3.4.3 Procedure ....................................................................................................................................... 32

3.4.4 Study System ................................................................................................................................... 33

3.4.5 Data ............................................................................................................................................... 34

3.4.6 Analysis .......................................................................................................................................... 36

3.5 Results ............................................................................................................................................. 36

3.5.1 Training Phase ................................................................................................................................. 37

3.5.2 Recall ............................................................................................................................................. 38

3.5.3 Learning Effects ............................................................................................................................... 41

3.5.4 Subjective Measures ......................................................................................................................... 43

3.5.5 Summary of Results ........................................................................................................................ 43

3.6 Discussion ....................................................................................................................................... 44

3.6.1 Why did Visual Terminal feedback outperform concurrent feedback schemes at training the task? ........................................................................................................................................... 44

3.6.2 Why did auditory perform better than other modalities for concurrent feedback? ........................................... 45

3.6.3 Why did Scale perform poorly? ...................................................................................................... 46
3.6.4 Why did the performance of concurrent visual feedback during training not transfer to recall? ................................................................. 47

3.6.5 Would the results translate to gait and other non-stationary situations? .... 47

3.7 Chapter Summary ............................................................................. 47

4 Multimodal Study .............................................................................. 49

4.1 Motivation ......................................................................................... 49

4.2 Experiment-Specific Background ...................................................... 49

4.2.1 Frequency of Feedback .................................................................. 49

4.2.2 Multiple Resource Theory .............................................................. 50

4.3 Experiment ......................................................................................... 50

4.3.1 Participants ..................................................................................... 51

4.3.2 Experimental Task .......................................................................... 52

4.3.3 Procedure ....................................................................................... 52

4.3.4 Study System ................................................................................ 52

4.3.5 Data ............................................................................................... 53

4.3.6 Analysis ........................................................................................ 53

4.4 Results ............................................................................................. 53

4.4.1 Training Phase ................................................................................ 54

4.4.2 Recall ............................................................................................. 54

4.4.3 Learning Effects ............................................................................. 57
4.4.4 Subjective Measures ............................................................... 59
4.4.5 Summary of Results ............................................................. 59
4.5 Discussion .............................................................................. 60
  4.5.1 Why was there no difference between the feedback schemes in recall load error? 60
  4.5.2 Why did Multimodal feedback outperform Visual Terminal during training? 61
4.6 Chapter Summary .................................................................. 61
5 Gait Study ................................................................................. 62
  5.1 Motivation ............................................................................ 62
  5.2 Experiment-Specific Background ........................................... 63
    5.2.1 Skill Transfer .................................................................. 63
    5.2.2 Steady-State Gait ............................................................ 64
  5.3 Experiment ............................................................................ 64
    5.3.1 Feedback Schemes .......................................................... 65
    5.3.2 Participants ...................................................................... 67
    5.3.3 Experimental Task .......................................................... 68
    5.3.4 Procedure ....................................................................... 68
    5.3.5 Study System ................................................................... 70
    5.3.6 Data ............................................................................... 71
5.3.7 Analysis.............................................................................................................. 71

5.4 Results ..................................................................................................................... 72

5.4.1 Training Phase ..................................................................................................... 73

5.4.2 Stationary Recall .................................................................................................. 73

5.4.3 Pre-Steady-State Gait Recall .............................................................................. 74

5.4.4 Steady-State Gait Recall ..................................................................................... 76

5.4.5 Learning Effects .................................................................................................. 80

5.4.6 Subjective Measures ......................................................................................... 81

5.4.7 Summary of Results .......................................................................................... 82

5.5 Discussion ............................................................................................................... 83

5.5.1 Why did stationary training methods outperform the gait training method
during training? ........................................................................................................... 83

5.5.2 Why was Visual Terminal more accurate than the other methods during
stationary recall? ......................................................................................................... 84

5.5.3 Why did Gait Summary training lead to underloading during recall? .............. 84

5.5.4 Why did Gait Summary training outperform Scale and Visual Terminal in
accuracy at the start of gait recall, but not during steady-state gait? ......................... 85

5.5.5 Why was the most precise feedback scheme during training (Scale) the least
precise during gait recall? Why was the least precise scheme during training (Gait
Summary) the most precise during gait recall? .......................................................... 86
List of Figures

Figure 2.1. A) Feedback control system [35] B) Feedforward control system [35]...... 13

Figure 3.1. Graphical comparison of feedback renderings throughout a loading attempt.
Each cell describes the feedback rendered by a given scheme at the moment indicated by
the vertical blue dashed line above it. The horizontal green line indicates the target load.
......................................................................................................................... 28

Figure 3.2. Experimental Setup. The cane device is attached to a laptop computer for
feedback. The Visual Full feedback scheme is displayed........................................ 33

Figure 3.3. A) Absolute mean loading error during training, aggregated by feedback
scheme. B) Absolute mean load error during recall, aggregated by feedback scheme. ... 39

Figure 3.4. A) Signed mean loading error during training, aggregated by feedback
scheme. B) Signed mean load error during recall, aggregated by feedback scheme........ 40

Figure 3.5. A) Standard deviation of loading error during training, aggregated by
feedback scheme. B) Standard deviation of load error during recall, aggregated by
feedback scheme. .................................................................................................. 41

Figure 3.6. A) Absolute mean loading error during recall, aggregated by session number.
B) Signed mean loading error during recall, aggregated by session number. C) Standard
deviation of loading error during recall, aggregated by session number. .................. 42

Figure 3.7. Total TLX (task load) score ± standard error for each feedback scheme.
Higher scores indicate greater workload for the feedback scheme; lower is better........ 43

Figure 4.1. A) Absolute mean load error during training, aggregated by feedback scheme.
B) Absolute mean load error during recall, aggregated by feedback scheme. ............. 55

x
Figure 4.2. A) Signed mean load error during training, aggregated by feedback scheme.
B) Signed mean load error during recall, aggregated by feedback scheme. .......................... 56

Figure 4.3. A) Standard deviation of load error during training, aggregated by feedback scheme. B) Standard deviation of load error during recall, aggregated by feedback scheme. ............................................................................................................ 57

Figure 4.4. A) Absolute mean load error during recall aggregated by session number. B) Signed mean load error during recall aggregated by session number. C) Standard deviation of load error during recall aggregated by session number. ............................................. 58

Figure 4.5. Total TLX (task load) score ± standard error for each feedback scheme. Higher scores indicate greater workload for the feedback scheme. Lower is better. ....... 59

Figure 5.1. Scale and Visual Terminal feedback screen on mobile device. For Scale feedback, the participant was shown the screen as the load number updated in real time. For Visual Terminal feedback, participants were shown a mean load after their attempt. The red box in the middle of the screen was a button for the experimenter to denote completion of a loading attempt. ............................................................................................................. 66

Figure 5.2. Gait feedback screen on mobile device. Each bar indicates the peak load through the cane for a given attempt, while the green line indicates the target load. Participants were shown the screen after each pass of the 12 ft course. ......................... 67

Figure 5.3. A) Absolute mean load error during the first 50 strides of gait recall, aggregated by feedback scheme. B) Signed mean load error during the first 50 strides of gait recall, aggregated by feedback scheme. C) Standard deviation of load error during the first 50 strides of gait recall, aggregated by feedback scheme. ............................................. 75
Figure 5.4. A) Absolute mean load error during training, aggregated by feedback scheme. B) Absolute mean load error during stationary recall, aggregated by feedback scheme. C) Absolute mean load error during steady-state gait recall, aggregated by feedback scheme.

Figure 5.5. A) Signed mean load error during training, aggregated by feedback scheme. B) Signed mean load error during stationary recall, aggregated by feedback scheme. C) Signed mean load error during steady-state gait recall, aggregated by feedback scheme.

Figure 5.6. A) Standard deviation of load error during training, aggregated by feedback scheme. B) Standard deviation of load error during stationary recall, aggregated by feedback scheme. C) Standard deviation of load error during steady-state gait recall, aggregated by feedback scheme.

Figure 5.7. A) Absolute mean load error during stationary recall, aggregated by session number. B) Absolute mean load error during steady-state gait recall, aggregated by session number. C) Signed mean load error during stationary recall aggregated by session number. D) Signed mean load error during steady-state gait recall aggregated by session number. E) Standard deviation of load error during stationary recall aggregated by session number. F) Standard deviation of load error during steady-state gait recall aggregated by session number.

Figure 5.8. Total TLX (task loading) score ± standard error for each feedback scheme. Higher scores indicate greater workload for the feedback scheme. Lower is better.

Figure A.1. Stationary Loading Experiment results by Target Load.

Figure A.2. Multimodal Experiment results by Target Load.
Figure A.3. Gait Experiment results by Target Load................................................................. 108
List of Symbols, Nomenclature or Abbreviations

PWB – Partial weight-bearing

BW – Body weight

DLS – Double leg support

SLS – Single leg support

COM – center of mass

EMG – Electromyography

TLX – Task Load Index

ANOVA – Analysis of variance

BLE – Bluetooth Low Energy

ADC – Analogue to digital converter

lb - Pound
1 Introduction

A variety of diseases, injuries, and surgeries can result in weakness of one or both lower extremities. Assistive devices, such as canes or crutches, are often prescribed to offload weight from an affected limb through the upper limbs. This rehabilitation intervention, called *partial weight-bearing* (PWB), aims to strike a balance in offloading weight from the affected limb: enough weight is needed to build strength during recovery, but not to the extent that further damage is sustained. The typical training approach used in clinical settings is to have people lean on a bathroom scale and transfer their weight to/from their assistive device until the scale reading matches their target load [20]. In an alternative training approach, a clinician may provide verbal feedback while the patient steps on the therapist's hand. Unfortunately, neither of these training techniques are overly effective, as evidenced by poor levels of compliance (i.e., accurately offloading a limb) [20]. Effective training in PWB is important since non-compliance can lead to complications and re-injury that are both expensive and reduce quality of life [8, 96]. Despite this, compliance is known to be poor with typical training methods, with less than 25% of cane users being able to offload within 25% of a trained target weight [20].

Compliance in PWB depends on the efficacy of this training as well as the user’s ability to transfer the learned skill to different contexts (e.g., walking and/or different assistive devices) [66]. The goal of training in such motor control tasks is to build an accurate *internal model* – a cognitive process that allows people to estimate and predict the outcomes of their actions [17]. Robust internal models allow skills to be better transferred to new situations. It has been shown that *augmented feedback* – providing added
information regarding task execution from external sources that wouldn’t be available to
the user through regular use – can play an integral role in the formation of these models,
and in learning of cause and effect relationships of the task \([45, 71, 75]\). Consequently,
the inadequate feedback given during conventional PWB training likely contributes to
poor compliance by limiting the user’s ability to build an accurate internal model \([84]\).

While research has explored how technology can be used to improve training, the effect
of alternative feedback is unclear due to incompatible testing conditions. For example,
recent work has found that training PWB during gait using specialized equipment (such
as instrumented assistive devices \([16, 30, 31]\), force plates \([20]\) or instrumented treadmills
\([52]\)), improves compliance by allowing training to occur in realistic, everyday
conditions. This work is limited, however, by only assessing functional PWB while
augmented feedback is applied. Furthermore, performing a skill while guided by
feedback is fundamentally different from learning the skill that the feedback aims to
teach. This is because regulating our actions according to guidance is not an effective
way to train a motor skill; one can follow guidance without learning how to do things
unassisted. It is therefore not clear if users of these systems learn the skill well enough for
long-term (post-feedback) compliance. In this work we constrain the training task to be
similar to current clinical methods without requiring expensive or complex equipment,
focusing on whether training with different feedback schemes may lead to improved post-
feedback compliance.

The remainder of this thesis will outline important background information regarding the
PWB task, present three experiments conducted to assess the effectiveness of PWB
training alternatives and discuss the implications of these findings in the context of PWB training and other motor tasks. Briefly here in summary, retrospective feedback was found to be better than concurrent feedback for training the PWB skill in a stationary context, though concurrent feedback outperformed retrospective at continuous skill regulation. This superiority, however, did not persist when the skill was transferred to gait. While gait-based training outperformed stationary training methods during the onset of gait, this advantage diminished over time post-training.
2 Background

This chapter discusses core information regarding gait, PWB, motor learning, and biofeedback. The first two sections explain why we need PWB, the third and fourth define PWB and how it is currently trained, the fifth outlines the psychomotor process of learning a motor skill, and the final section describes biofeedback and how it has been used for PWB training. This knowledge is critical for contextualizing the problem addressed by this work: how to use biofeedback to improve PWB training.

2.1 Mechanics of Gait

Mobility is a crucial aspect of maintaining functional independence, especially in senior populations [57]. Humans have evolved to ambulate through bipedal locomotion. The simplest physical model for this process is an inverted pendulum chain [58]. In this model, the ground acts as an anchor while a pendulum is in contact with it and rotation about this point of contact features a translational component in the intended direction of travel. Torque about the center of the chain then rotates the other pendulum leg forward to anchor on the ground at a position further in the direction of travel than the previous anchor point. This alternating sequence of steps continues throughout the activity of gait. Human anatomy is notably more complex than this simple model, with multiple joints in the lower extremities coordinating to produce this gait pattern. The basic concepts of torque and anchorage still apply though.

A variety of skeletal muscles span the lower extremities, anchored to adjacent bones at a displacement from their connecting joints. Contraction of a muscle, therefore, produces a
torque through the joint it spans [23]. These torques allow rigid segments such as the foot and thigh to rotate and translate relative to each other. Gait requires translation relative to the ground, necessitating external forces. Vertical forces are mostly canceled out; gravity is countered by normal forces through the feet while they contact the ground. Horizontal frictional forces are also produced to resist foot slippage as the rest of the body moves forward relative to it [33]. External forces are the basis of locomotion, but it is often more insightful to view body movement holistically.

On a higher level, there are common physical features that can be used to describe an individual’s gait progression. A gait cycle is typically defined as commencing when a heel strikes the ground and concluding with the subsequent strike of the same heel. Within this cycle, phases can be extracted. The simplest segmentation scheme features “contact” while a given leg is bearing weight and “swing” throughout the rest of the cycle. Subdividing the contact phase provides a more comprehensive model that considers the contralateral (i.e. opposite) leg’s actions throughout the gait cycle. Heel strike initiates “double leg support” (DLS) where both legs are in contact with the ground. This is followed by the contralateral leg rising from the ground, leaving a single leg to support the body. This “single leg support” (SLS) phase concludes with a contralateral heel strike, initiating a second DLS phase. The ipsilateral (i.e. original) leg then rises into the air for its swing phase [79]. These phases can be further subdivided to examine the gait cycle’s finer details, but such endeavours are beyond the scope of this thesis.
Each gait phase is a sequence of coordinated movements of the multiple rigid segments that comprise the human body. These movements are combined in ways that minimize energy expenditure. Simple bipedal locomotion induces periodic vertical displacement of the body’s center of mass (COM). Raising the COM expends an appreciable amount of energy which is converted to kinetic energy as the COM lowers. Ideally this kinetic energy would then be harnessed to lift the COM again, but energy is typically lost (e.g., to sound, heat, vibration, etc.) in the conversion process. Movements of the foot, ankle and knee can help improve the efficiency of this transfer by smoothing the gait cycle [33]. Humans have also adapted gait strategies to reduce the magnitude of energy transfer by minimizing the COM’s oscillation amplitude [33]. Hip rotation in the transverse (i.e. separating the body into top and bottom sections) and coronal (i.e. separating the body into front and back sections) planes destructively interferes with the COM oscillation created by bipedal locomotion, reducing its amplitude and increasing forward displacement per stride [33]. Transverse plane rotation can become unstable at higher gait speeds but is effectively dampened by countering with rotation of the upper body [33]. Differences in segment proportions, flexibility, and relative muscle strength lead individuals to adopt their own personalized strategies for minimizing energy expenditure during gait [33].
2.2 Gait monitoring methods

Since the variability in gait patterns between individuals can be large, identifying deficiencies in one’s gait can be challenging. It requires a detailed understanding of the underlying physical factors influencing an individual’s gait. Clinical gait monitoring produces objective evidence for characterizing these factors. Assessment is typically performed by a gait specialist using one or more of a variety of data collection methods. Method selection depends on both the specialist’s expertise and the scope of assessment.

Figure 2.1. Anatomical planes for division of the human body [72].
Electromyography (EMG), the study of electrical signals generated during the contraction of a muscle, was one of the earliest objective tools for gait analysis. Electrodes are generally secured over pertinent lower extremity muscles while the patient walks. A muscle is considered active during intervals when the magnitude of the EMG signal exceeds an activation threshold. Temporal and strength abnormalities can be identified by comparing the activation series of different muscles. Very specific information can be gleaned from this method, but the specialization of expertise and equipment required to execute EMG assessment makes it rather uncommon [60].

Clinical video analysis is a simpler alternative, requiring only a video camera. Clinicians align the camera such that it captures either the patient’s coronal (from the front) or sagittal plane (from the side) of motion while they walk. With a fixed camera position, the patient is only truly viewed in sagittal plane at the center of the image, restricting the usable observation to a few strides. After observation, however, video frames can be annotated with joint angles and relative segment positions. These measurements can be more accurately obtained using more formal motion capture analysis for a comprehensive assessment of joint kinematics throughout gait [2].

Almost any clinical gait assessment includes a physical examination because standardized kinematic and kinetic parameters can readily be translated between clinicians with little clarification required. Patients lie or sit in various positions as the therapist manipulates their lower extremities to simulate motions relevant to gait. Maximum flexion and extension angles for each limb are measured, along with the lengths of various segments. Deformities in bones or joints are also noted. A 6-point test
of strength is also conducted for pertinent muscle groups, quantifying their ability to resist gravity and/or forces applied by the therapist [36]. This examination allows clinicians to identify physical limitations of a person’s body that may affect gait.

A variety of quantifiable measures can be extracted from clinical gait analysis by combining these various approaches. Gait speed, cadence, and distance per stride can be calculated from video and/or EMG. A person’s kinematic and kinetic capacities are gleaned from physical assessment. Energy expenditure may also be examined by monitoring oxygen consumption during ambulation. It is often beneficial to compare findings to a baseline template, facilitated by the normalization of kinetic (by weight) and kinematic (by height or length) parameters. Defining regular gait, however, remains difficult due to interpersonal variations in gait strategies even among healthy individuals [2]. Despite this issue, gait analysis is a common tool for identifying the physical mechanisms that contribute to gait deficits. Such findings are useful for informing surgeries and interventions.

2.3 Partial Weight-Bearing

As previously discussed, gravitational forces on the body during gait are counteracted by reactionary forces on the feet. These forces propagate vertically through the body, with each successive section supporting the weight of segments above it. Consequently, the weight-bearing capacity of a lower extremity is limited by its weakest component. Joints and segments can be compromised by injury, degeneration, or surgery. Fractures of the femur or hip are especially common in senior populations, compromising the integrity of
the affected bone [1, 56]. Increasingly common surgeries such as total hip arthroplasty also disrupt bone integrity [27].

Post-operative (or post-injury) rehabilitation aims to strengthen the compromised joint or segment, returning the overall limb to its prior weight-bearing capacity [29]. Bones have been shown to adapt to environmental conditions, requiring some degree of weight-bearing stress to stimulate growth [29]. However, overloading a compromised bone can delay healing or worsen the damage [10, 29]. Rehabilitation of the lower extremity requires a balance between protection and growth stimulation. Therapists and/or surgeons prescribe specific weight-bearing limitations on a limb during rehabilitation to optimize both factors. There are three common levels of weight-bearing prescription: toe-touch, partial weight-bearing (PWB), and weight-bearing as tolerated. Toe-touch limits weight-bearing through the affected limb to less than 20% of the patient’s body weight (BW). Partial weight-bearing allows for more weight to be borne by the affected limb with prescribed targets often ranging between 20% and 50% of BW. These definitions are a rough approximation, with many medical professionals settling on slightly different boundaries [22, 29]. Weight-bearing as tolerated allows the patient to select a weight-bearing level based on the pain it induces in the limb. Ultimately, patients are often prescribed a progressive weight-bearing program, increasing the load through their affected limb as it heals.

Limiting the weight borne by a limb has applications beyond bone-based injuries. Matrix-induced autologous chondrocyte implantation and osteochondral autograft transplant surgeries repair knee cartilage. In both cases, progressive PWB rehabilitation improves
functional ability of the knee post-surgery [44, 80]. PWB also improves functional ability when the joint is affected by hemophilic knee arthritis and can slow the progression of knee osteoarthritis [61, 95]. Both aforementioned surgeries are interventions for knee osteoarthritis, so partial weight-bearing is particularly critical for people living with it. Rehabilitation from other pathologies, such as stroke, also frequently involves PWB to improve gait quality [5]. Even when it isn’t used as a therapeutic tool, weight-bearing ability is a key component of symmetry assessment for neurological disorders [25], and stability in seniors.

2.4 PWB Training

Weight-bearing can be shifted to the unaffected limb during the DLS phases of gait, but the affected limb bears the full brunt of body support during its SLS phase. Assistive devices such as canes, crutches, or walkers are utilized as (an) additional point(s) of contact to share this load [20]. A percentage of the total SLS load can be transferred through these devices to the upper extremities, allowing the affected lower extremity to bear appropriate loads that both protect it and promote healing. Patients may be verbally instructed to offload a target load to the assistive device, but the actual loads produced have been shown to vary drastically from these instructions [31, 61].

There are currently two methods commonly used in clinical practice for training PWB: the scale and the hand-under-foot methods. Both methods feature patients in a stationary stance with both feet and their assistive device planted. For the scale method, the affected limb rests on a bathroom scale while the unaffected limb is planted on a surface raised to the height of the scale. The patient transfers weight between their assistive device and
affected limb until the scale reading matches the target set by their therapist. For the hand-under-foot method, the therapist places their hand (or foot) under the patient’s affected limb. The therapist verbally instructs the patient to increase or decrease the load through the limb until they deem the force to match their target [20].

Despite their widespread use, conventional training methods have consistently been shown to be ineffective at teaching people to accurately comply with PWB targets. The hand-under-foot method relies on subjective and imprecise estimation by the therapist; their perception of load through the affected limb may be off by up to 30% [29], whereas the scale method features more accurate load measurement. Patients taught by this method can reliably offload their affected limb to a target level while stationary [29]. Despite this, neither method is effective at training people to accurately offload their limb to a target level during gait [20, 64]. Patients learning PWB through these training methods often overload the limb, putting them at risk of further injury or slowed recovery [29].

2.5 Internal Models

Execution of a motor task is a complex process that is often taken for granted. The simple act of reaching, for example, requires coordinated activation of multiple muscle groups controlling the shoulder, elbow, wrist and hand. Using senses such as sight and proprioception, muscle activations are honed until the limb assumes the desired orientation. This can be modelled as a feedback control system, shown in Figure 2.2A. The cerebellum acts as a controller, sending activation signals to coordinate muscular contractions. Muscles act as plants, receiving these signals and performing the
corresponding actions. Limb positions are sensed by vision and proprioception and relayed back to the cerebellum. This information is used to update the controller’s perception of limb position and inform the next iteration of activation signals to bring the system closer to its goal [35].

Figure 2.2. A) Feedback control system. [35] B) Feedforward control system [35]

The controller must implicitly understand how the plant will react to stimulus, and this “internal model” of the plant can be abstracted out into a separate component. Feedforward models fork signals from the simplified controller to pass through both the plant and internal model (Figure 2.2B). Outcomes predicted by the internal model are corroborated by sensory feedback of the plant’s actual actions and returned to the controller. A well trained feedforward system can still operate in the absence of sensory feedback by iteratively acting on the plant outcomes predicted by its internal model [35]. Feedback is rarely entirely absent or isolated to one modality, however; humans have a wealth of senses at their disposal and may utilize visual, auditory, or haptic feedback to
complete motor tasks. Haptic feedback is especially important to such tasks, encompassing the body’s kinematic and kinetic states [62].

Feedback can be further classified by the systems from which it initiates. Intrinsic feedback relates to a system’s perception of its plant’s actions. In the context of PWB, patients use intrinsic proprioceptive feedback to estimate the load through their limb or assistive device. External sources of feedback, such as scale readings or verbal commands from a therapist, constitute extrinsic feedback [83]. These PWB training methods improve a patient’s understanding of the task performance measure: weight-bearing. Their internal model uses this extrinsic feedback to adapt their mapping of muscle stimuli and/or proprioceptive feedback to load on the affected limb [35]. Conclusion of training severs this adaptive extrinsic process, leaving the patient to recall PWB loads from a feedforward system with intrinsic feedback.

Changes in the environment can be addressed using feedback, but it is more difficult to react to changes in the motor task itself. Fatigued muscles have reduced maximum force capacities, requiring more stimulation than rested muscles to produce the same level of force. Humans struggle to adapt to this factor, underestimating kinetic requirements when a limb is fatigued [77], attempting to apply an internal model developed using one set of states/parameters to a modified task (i.e., differing fatigue levels). This is an example of poor skill transfer. Conversely, a professional squash player may find that they also excel at badminton because their racquet stroke skills transfer more readily. There is still a learning curve, however; the player may, for example, initiate their first few swings of the lighter racquet too early, missing the shuttle entirely. A similar effect can be seen
with current clinical PWB training techniques. Patients learn the skill while standing stationary, then try to apply this intrinsic feedforward control loop to recall their PWB target during gait. This is a drastically different weight-bearing environment, with load on the affected limb exceeding the total stationary load by 20% during the initiation and termination of its SLS phase [33]. Current PWB training methods adapt a patient’s internal model of the stationary task, but the patient must then transfer this skill to a different task. The scale method has been shown effective at teaching stationary PWB, so skill transfer is likely a primary cause of poor PWB compliance during gait [20, 29].

2.6 Guidance

Extrinsic feedback is effective at guiding users toward proper task execution during training, improving performance while the feedback is present. In some cases, this may be beneficial. Complex motor tasks feature a large number of parameters that must be coordinated spatially and temporally, making it difficult for novices to execute them properly. Concurrent augmented feedback can help users discern the impact of each degree of freedom on the overall task, allowing them to hone in on and experience proper task execution [75]. Effectively, adding complementary feedback can make the task easier to execute, improving performance while feedback is given.

This added guidance is not, however, most beneficial in all cases. Psychology research into motor and verbal tasks has long found that introducing difficulty and variability into training better improves skill retention and transfer [66]. Its benefits are known to expand into spatial memory as well [11]. A simple yet effective way of making practice more difficult is withholding feedback (i.e., only offering it after a certain number of training
If feedback is too frequent it may lead users to rely on it for regulation, preventing introspection and proprioceptive comprehension which is critical to unassisted motor task execution [66]. The “guidance” hypothesis indicates that frequent augmented feedback becomes embedded as part of the internal model of the motor task, degrading performance when it is removed during later task execution [59, 68, 69]. As the most frequent type of feedback available, concurrent feedback could therefore be expected to limit the benefits of training. Conversely, a try-and-test style of retrospective feedback precludes real-time regulation, forcing users to actively develop a reliable internal model without the aid of the added feedback [74, 84]. This may lend itself to superior retention of simple motor tasks where there are limited degrees of freedom to comprehend, but may also limit the initial learning process. While previous research has shown mixed results on the effects of delay [66], it is known in general that low latency (concurrent) feedback is more effective at training complex tasks, whereas higher-latency (terminal) schemes are better for simple tasks [74, 75].

For the remainder of this thesis, the term “guidance” is used to define extrinsic feedback guiding a user to the target while it is applied. This is largely viewed as negative for motor learning because users come to rely on this extrinsic feedback, inhibiting learning and impairing performance when feedback is removed.

2.7 Biofeedback for PWB

A logical way to improve PWB training is to eliminate the skill transfer step by training patients while they ambulate. Force plates can capture tangential and normal forces applied to them with high degrees of accuracy and precision, often being used as a gold
standard to compare pressure- or force-sensitive products against [34, 38, 48]. These electronic devices can also sample at high rates, capturing detailed time series of applied forces as a user steps across them. Summary feedback from force plates after gait therefore drastically outperforms the scale and hand-under-foot methods of teaching PWB for target recall during gait [20]. Unfortunately, these devices are incredibly expensive to obtain and maintain, so it is infeasible to install more than a few in a laboratory and is prohibitive, clinically. Furthermore, only one or two gait cycles can be captured over such short distances, so they are unable to assess PWB loads for prolonged gait. Despite these limitations, force plates have become the gold standard for calibrating more economical and mobile PWB monitoring systems [26].

Recent developments in embedded system technology have made it feasible to instrument passive devices to monitor their kinetics and kinematics. When these “smart” devices are worn or held by humans, they can be used to assess human movement by proxy. Smart insoles can be placed in the shoe of an affected limb to accurately monitor the load it is bearing [3]. These devices can continuously monitor a patient’s PWB compliance, with wireless models promising the potential for outpatient deployment in any environment [21]. Commercially available models offer audio or vibrotactile feedback when wearers exceed a threshold load through the insole. Patients trained to PWB in this manner outperform those trained by the current clinical methods and retain their PWB compliance days after feedback is removed [22, 28, 32]. The sensors in currently available systems degrade quickly, however, and their cost of replacement makes reliable long-term load monitoring unfeasible [38].
Partial weight-bearing transfers weight from an affected limb to an assistive device; as such, these devices can also be equipped as an alternative tool for monitoring PWB compliance. Whereas insoles must be customized, fit, and installed in a patient’s shoe (where size, heat and sweat are major confounding factors), instrumented assistive devices such as a smart cane, can be readily adopted without modification. Canes have been popular devices for PWB biofeedback, with various models having been used in research for decades [14]. Assistive devices lend themselves to prototyping, as even bulky additions have minimal impact on a person’s assisted gait pattern. A variety of smart canes have been developed and tested as PWB training devices, with each utilizing either audio [37, 49], vibrotactile [16, 61], or visual [87] feedback. Audio and vibrotactile modalities are conducive to real-time interventions, while visual feedback is typically provided as a performance summary after a group of gait cycles. Visual feedback is also sometimes provided to the attending therapist so that they can monitor their patient’s performance [14]. Like smart insoles, these canes outperform current clinical methods at PWB training [29, 37, 52]. Similar studies using vibrotactile feedback during crutch-assisted PWB have found that it too outperforms current clinical training techniques [16].

There is a substantial body of evidence supporting the benefit of device-based biofeedback for PWB training. It is difficult, however, to directly compare the effectiveness of each, given differences in experimental conditions and methodologies. Feedback, for example, may be discrete [20] or continuously rendered while a condition is met [61], whereas some devices use bandwidth feedback, treating the target as a range of acceptable values rather than one discrete value [14]. Each device considered in these
studies has been customized to provide one feedback modality, each of which has several other considerations. Audio and vibrotactile schemes, for example, can vary in both frequency and intensity. The human body is not uniformly sensitive to vibrotactile stimulation, so placement of vibrotactors (which provide vibrotactile feedback) is also an important consideration [12]. Visual feedback, in summary form, requires visualization design choices, a field of its own with numerous considerations [51]. Importantly, these factors cover feedback design for a variety of PWB training devices, with the effectiveness of each training scheme having been assessed on devices of varying configuration and accuracy, and with different PWB targets. Even success metrics varied; some studies quantified success as the percentage of strides within an acceptable target load range [7], while others tracked differences between actual loads and discrete targets. Furthermore, some based their feedback on leg load whereas others used device load. While the distinction may seem trivial, the approaches may not be related linearly, due to the fluctuation of borne weight throughout the gait cycle [33]. Apart from a few studies with commercially available smart insoles, controlled comparisons of PWB training performance have not been conducted for these biofeedback schemes. The outcomes of the few studies that have compared insole devices have been limited to usability surveys and accuracy of load measurements [43, 48].

Retention of PWB compliance post-feedback, a critical measure of how well a feedback method trains the motor skill, has also only been tested in a few studies [32, 61, 84]. To identify an optimal training method for this skill, diverse feedback alternatives must be administered and evaluated under comparable conditions. Despite its importance for
improving rehabilitation outcomes, previous research has not compared the effectiveness of different feedback permutations for PWB training.

This thesis presents three experiments to address this gap in knowledge. The first compares the effectiveness of stationary PWB training methods. The second considers whether performance can be improved by combining two different training methods. Finally, stationary and gait training methods are assessed during both stationary and gait recall to determine how best to train the skill for each context.
3 Stationary Loading Study

In this chapter we investigate optimization of stationary PWB training by studying the performance of various feedback alternatives. The following sections will outline the motivations behind this study, describe the experiment, present its results, and discuss the implications of these results along with potential future work.

3.1 Motivation

Current clinical methods have repeatedly been shown to be less effective at training PWB gait than gait-based biofeedback devices [16, 20, 29, 30]. However, the primary training method used in clinical settings still teaches users partial weight-bearing while stationary since this is the easiest way to provide feedback (e.g., by using a scale). Previous research has evaluated static training methods, but then required participants to transfer their newly learned stationary skill to a new, unpracticed task: gait. Transferring skills to a new scenario or task is well-known to harm performance because users must extrapolate their cause-and-effect understanding of the task to incorporate new objectives or external factors [9, 66].

While gait-based training is superior for training gait PWB skills, stationary training has several advantages (such as less or no complex hardware, and that it is the current clinical standard) and so assessing how it can be improved is worthwhile. Thus, optimizing current stationary PWB training methods could improve PWB compliance through skill acquisition (decreasing overall error, even during gait) and could likely be more readily adopted by clinicians. For these reasons, this first experiment foregoes the skill transfer
step and specifically targets stationary PWB performance. By training and testing the skill during the same task, the impact of feedback style on training effectiveness can be isolated to identify superior stationary training methods. Further, our understanding of motor learning suggests that with improved internal models, performance when transferring the model to new task scenarios will be improved [39, 68].

3.2 Experiment-Specific Background

3.2.1 Feedforward Control in Partial Weight-Bearing

Determining which feedback is best for a task requires some consideration of how the feedback is used. There are two main use cases for this:

1. **Regulation**: Feedback that is always present to assist a user in regulating the task’s performance measure.

2. **Retention**: Feedback that is briefly used to teach a user how to properly perform the task, allowing the user to accurately perform it later without augmented feedback.

Both use cases (continuous feedback or training only) are applicable to partial weight-bearing and can be modelled as feedforward control processes. In feedforward motor control tasks, the brain acts as a controller for an aggregation of muscles that facilitate movement of the limbs [50]. In PWB, the movement of concern is how these muscles distribute weight throughout the musculoskeletal system. During the single-leg stance phase of gait, load is balanced between two points of ground contact: the weakened leg and an assistive device (or devices, in the case of crutches) [20]. Based on proprioceptive feedback from their leg and supportive arm(s), users can estimate how much load is
going through each of these supportive limbs and adjust their loading accordingly [46]. In the absence of damage to the nervous system, this intrinsic feedback (i.e., feedback created by the user) is always present. However, mapping proprioceptive “feelings” to quantifiable loads is not necessarily intuitive; augmented extrinsic feedback (i.e., created by an external source) can help a user develop better internal models [74]. During PWB training, users are offered both augmented (extrinsic) and proprioceptive (intrinsic) feedback to inform the brain of the balance adjustments necessary to achieve their desired loading pattern. Unless otherwise specified, the term “feedback” in this thesis will refer to augmented feedback. Extrinsic feedback simultaneously helps users regulate balance and build an internal model that maps proprioception to actual load. After training, augmented feedback assistance is typically removed, and users must rely solely on their intrinsic perceptions and internal model of the PWB task. While systems can also be designed for continuous use (i.e., feedback is always provided), this can cause a user to develop an internal model that is reliant on the feedback; see the Guidance section of the Background chapter for more details [55, 74]. These two use cases have different performance requirements: continuous regulation systems need only improve PWB compliance while feedback is applied, whereas conventional training methods focus on compliance after feedback has been removed. Feedback schemes should be optimized for their intended use case, but the effects of extrinsic feedback parameters on PWB compliance during and after application are not well understood.
3.2.2 Feedback for PWB Training

Designing the optimal feedback scheme for a training task is multi-factorial, and basic questions include: Delay: When should feedback be provided (i.e., concurrently–during the task–or, terminally–after the task)? Resolution: What level of detail would be most effective (e.g., fine-grained or coarse-grained feedback)? Modality: What modality would be most effective (i.e., haptic, auditory, or visual)? Each of these question areas is elaborated on below.

Feedback delay relates to when the feedback is provided, either while the task is being learned or afterwards, in summary (i.e., upon completion of a task or set of tasks). Concurrent feedback provided during a motor control task may allow a learner to understand and make adjustments to their control rapidly, allowing them to more easily train their internal model. However, concurrent feedback might also distract the learner or train them to only use the extrinsic feedback, meaning they may lose the performance benefits when this feedback is removed. Terminal feedback (provided after the task is complete) would encourage the learner to reflect on their actions and how they might be adjusted, explicitly and deliberately training their internal model. However, terminal feedback may also provide information that is too complex, removed from the original task, or abstract for a learner to understand how their actions can be successfully adjusted.

Feedback resolution describes the number of levels that can be rendered. The lowest resolution is binary, typically an alarm that is triggered when a certain condition is met. For example, auditory alarms have been shown to improve acquisition and retention of
proper hip flexion in gymnasts performing circles on a pommel horse [4]. Some feedback methods use a tiered approach with a small number of discrete levels [14]. Higher resolution schemes feature continuous gradients of output, such as modulation of frequency or intensity of vibration or sound [42]. It is common to combine continuous and tiered approaches to create a form of bandwidth feedback: continuous output when the performance variable is outside a set range of values, and constant output (often silence or no vibration) while it is within the range. Resolution has a similar effect to feedback delay: low-resolution feedback can obstruct the user’s knowledge of their current performance to make them rely more on their intrinsic feedback. Although high-resolution schemes facilitate finer adjustments, they may result in dependence on the augmented feedback [42, 63].

Feedback modality considers which sense is targeted. Different modalities of feedback have been used for a variety of training tasks, from which some generalizations can be drawn [15, 47, 59, 75]. Visual feedback is often more effective than other modalities at training complex spatial motor tasks, likely due to the acuity and bandwidth of human vision relative to the other senses [74]. Home therapy routines have used visual feedback from Microsoft Kinects to improve user accuracy in following prescribed exercises [92]. Conversely, auditory feedback has frequently been shown to be optimal for rhythmic training tasks [59, 73]. Haptic feedback has also been frequently investigated, likely because of the perceived similarity between it and intrinsic feedback in PWB: a combination of touch and perceived load on the limbs. However, haptic feedback has
received mixed results, both when applied in isolation (unimodal) and in systems that combine it with audio or visual feedback (multimodal) [15, 47].

The selection of an appropriate augmented feedback modality for internal model development therefore requires classification of the motor task as being either simple or complex; see the Guidance section of the Background chapter for more details. The complexity of a task impacts both when the feedback should be given and which feedback modality should be used. The complexity of PWB is not obvious: although it requires the coordination of multiple muscle groups, its performance assessment is often based on a unidimensional quantity (load). This ambiguity makes it important to consider feedback alternatives that could favour either complexity.

3.2.3 Feedback Design for Training Motor Control Tasks

Some work has previously touched on aspects of designing augmented assistance for motor control tasks to develop improved mental models. For example, research has been conducted on motor skill acquisition in video games [24], the design of visual feedback in rehabilitation for stroke patients [65], general motion guidance using visual, haptic and spatial audio feedback [70], and visual guidance for complex 3D gestures [13, 78]. However, little work has compared different modalities or made specific considerations for the design of feedback for motor control learning. Given that the effectiveness of different modalities varies depending on task characteristics [74], any existing work from other fields may not directly apply to PWB.
3.3 Feedback Design

The three characteristics of feedback (modality, resolution, and delay) are universally important considerations when designing motor task training systems. They are independent of task performance criteria, desired performance level, training schedule, and other user-specific parameters.

Here, we designed 6 feedback schemes to cover the most common and feasible permutations of the characteristics outlined in Feedback for PWB Training, along with a “Scale” scheme that mimics current clinical training practices for PWB; see Figure 3.1. Every permutation of modality and resolution was tested, with the exception of continuous-resolution haptic feedback, which scored poorly in pilot testing (it was found to be difficult to perceive near the target).

In order to isolate the benefits of each modality, we focused on unimodal applications rather than multimodal (i.e., the use of multiple modalities together at once). Further, certain modalities were omitted due to practical limitations of the clinical environment (e.g., visual feedback requires a display).

Concurrent feedback schemes were emphasized because they have been shown to be effective at PWB regulation in past research [16, 49, 61], but their capacity for building internal models has not been previously assessed. One terminal scheme was developed to assess the effect of delaying feedback, a characteristic that is known to build more robust internal models than concurrent feedback in some tasks [55, 75].
Figure 3.1. Graphical comparison of feedback renderings throughout a loading attempt. Each cell describes the feedback rendered by a given scheme at the moment indicated by the vertical blue dashed line above it. The horizontal green line indicates the target load.
3.3.1 Feedback Schemes

In total, 7 feedback schemes were developed and tested, as summarized below and in Figure 3.1:

1. **Scale** – visual, continuous, concurrent feedback. Displays a numeric value on the screen with one decimal point of precision to indicate the current load on the cane. Feedback was provided at 10 Hz for readability while maintaining responsiveness.

2. **Visual Full** – visual, continuous, concurrent feedback. Displays a red vertical bar with a small green “success zone” in the center. This zone represents 2 lb above and below the target. A white arrow moves along the bar to indicate where the current cane load is in relation to the target, with overloading being higher than the success zone. An additional check mark lights up in the success zone while the arrow is within it.

3. **Visual Bandwidth** – visual, bandwidth, concurrent feedback. Identical to Visual Full, but the arrow disappears within the success zone.

4. **Audio Full** – audio, continuous, concurrent feedback. Plays a 493.88 Hz tone (B4 note) when the load is below the target and a 987.77 Hz tone (B5 note) when it is above it. Intensity decreases as load approaches the target, and the waveform shifts from a harsh sawtooth wave to a softer sine wave when load is within the success zone.

5. **Audio Bandwidth** – audio, bandwidth, concurrent feedback. Identical to Audio Full, but the tone reduces to silence in the success zone.
6. **Haptic Bandwidth** – haptic, bandwidth, concurrent feedback. This haptic equivalent of Audio Bandwidth features a vibrational motor worn on the participant’s wrist. There is no high/low frequency distinction, only intensity indicating discrepancy between load and success zone.

7. **Visual Terminal** – visual, continuous, terminal feedback. Features a numeric display like the Scale method but provides no feedback during the attempt. Once the cane has been released, the display is updated to reflect the mean load held for 1 second after the participant indicates they have reached the target (See Procedure for more details).

### 3.4 Experiment

The goal of this study was to evaluate the performance of 7 combinations of the three characteristics: modality (3 levels: audio, visual or haptic), delay (2 levels: concurrent or terminal), and precision (2 levels: bandwidth or full). While these factors suggested up to 12 combinations of the 3 factors (3x2x2), we initially identified 7 schemes that provided the most logical combination of the factors (as described above). The Scale method was chosen because it represents the current clinical practice. A terminal method was developed to mimic the Scale method, but with altered delay. Only one terminal method was developed because most experimental systems in the literature offered concurrent feedback, so it seemed preferable to more fully compare the alternatives in this space. The remaining methods cover two levels of precision for each modality except haptic, which was determined too difficult to perceive at lower intensities in piloting. Because
factors were not fully crossed, evaluation of feedback was treated as a single factor with 7 levels.

The study consisted of seven sessions, each on different days. Each session tested one feedback scheme and lasted 15-30 minutes. Sessions were conducted on different days to minimize the learning effect and participant fatigue. Sessions were not required to occur on consecutive days, nor were they required to be conducted at the same time of day. Each participant received a $10 honorarium.

3.4.1 Participants

Fourteen healthy university students and staff (7 identified as female and 7 as male; mean age: 27 years; min: 20; max: 35; s.d.: 4) volunteered for the experiment, which was advertised through a campus email list and posted fliers. Exclusion criteria were an inability to lean body weight on one leg, an inability to support 20% of body weight through dominant arm, or more than 5 days elapsing between subsequent sessions; one participant was excluded for a gap of 6 days between sessions.

Demographic information such as age and dominant hand was collected at the beginning of the first session, along with body weight and height for calibrating the target loads and cane height respectively. One participant reported being left hand dominant, with the remaining 13 being right hand dominant. Mean body weight was 158 lb with a minimum of 116 lb, a maximum of 205 lb, and 24 lb standard deviation. Five participants reported having no prior experience using walking assist devices, 4 reported less than one week of
experience, 2 reported between one week and one month, 2 reported between one month and one year, and 1 reported more than one year of experience.

### 3.4.2 Experimental Task

The primary task was designed to mimic the protocol of clinical partial weight-bearing training sessions. The primary task of each session was loading an instrumented cane with the dominant arm while standing stationary. Participants would train with the session’s feedback scheme to load the cane to a defined target load. They were then distracted to clear their short-term memory, and subsequently asked to recall the target load without assistance.

### 3.4.3 Procedure

During the first session, participants were explained the experimental procedure, provided informed consent, and completed a demographic questionnaire. They were also weighed on an analog bathroom scale to allow for proper cane calibration.

Each session featured one feedback scheme for training. The cane height setting was sized to the participant according to standard clinical practice at the beginning of each session [41]. The session began with a training phase: the participant would lean on the cane while receiving feedback assistance to help them achieve one of four target loads within the clinically loading range of a cane (8%, 10%, 12%, or 14% of their body weight). Participants were informed that they would later need to recall this load without assistance. They would use the feedback assistance to make 5 attempts at loading to the target, verbally notifying the experimenter when the target was achieved and releasing
the cane to rub their hands together between attempts. This was done to reduce any short-term haptic memory of the loading attempt from the hand surface [19]. After these 5 training attempts, participants would be distracted by playing 1 minute of a video game based on the Stroop Test [53]. This distraction phase was followed by a recall phase: 5 attempts at loading the cane to the target without feedback assistance. Training, distraction, and recall phases were then repeated for the other three target loads. Target load and feedback orderings were pseudo-randomized for balance across the 14 participants.

![Experimental Setup](image)

**Figure 3.2.** Experimental Setup. The cane device is attached to a laptop computer for feedback. The Visual Full feedback scheme is displayed.

### 3.4.4 Study System

In order to evaluate the feedback schemes, a stationary “smart” cane system was developed. The “smart” cane for this experiment is a regular swan neck (also termed “offset”) walking cane with a Wheatstone bridge of strain gauges in the neck of the device. This smart cane is a simplified version of the one described in [18], with an Arduino Uno attached to the strain gauge. Output voltages are discretized by the
Arduino’s 10-bit ADC (Analog-to-digital converter) and passed via USB to a connected laptop computer; see Figure 3.2. An application running on the laptop (written in Processing) reads in the ADC value and maps it to a load in pounds. The mapping uses a calibration curve (collected at 5 lb increments) to interpolate actual cane load from ADC value. Calibration was performed up to 100 lbs, and there was no evidence of hysteresis or drift within this range. However, even with care taken in grip consistency, the actual load for each sensor reading could vary by ±5 lb between calibration attempts.

A sampling rate of 30 Hz allowed loading data to be regularly and reliably collected by the application. Each load reading was logged to a text file along with the timestamp, target load, anonymized participant ID, feedback condition, and experimental phase (training, recall). If the participant was undergoing a training phase, the loading value was also used to render feedback.

3.4.5 Data

In order to evaluate the effectiveness of our feedback schemes, the following three dependent variables were calculated: absolute mean load error, signed mean load error and standard deviation of load error. All three measures are standard instruments for quantifying performance in motor control tasks [67].

Load, in the context of this study, is defined as the magnitude of the normal force exerted by the participant on the cane handle (parallel to the cane shaft). Pounds (lb) refer to pound-forces, equivalent to the magnitude of the force exerted by gravity on a 1 lb mass. Body weight is the magnitude of the net force exerted by gravity on a participant’s body.
Load error for five trials (for each target load and feedback scheme combination) was quantified as the difference between the target load and the mean of load samples for 1 second (i.e., 30 samples) after the participant indicated that they had reached the target load. Mean and standard deviation values were calculated across the 5 attempts for each condition. Load error could be negative (indicating under-loading) or positive (indicating overloading); a mean of zero indicated perfect loading.

Accuracy: Absolute mean load error quantifies the accuracy of PWB, indicating how close to the actual target the user is (often referred to as absolute constant error [67]), calculated as: $|\sum(x_i - T)/n|$, where $T$ represents the target load and $x_i$ represents the load achieved on a single trial.

Signed Accuracy: Signed mean load error quantifies both how close to the actual target the user is, and the direction of their error (i.e., over or under target): $\sum(x_i - T)/n$, where $T$ represents the target load and $x_i$ represents the load achieved on a single trial. This measure is often referred to as constant error [67].

Precision: Standard deviation of load error quantifies the precision of PWB, indicating how consistent loading is to a perceived target (often referred to as variable error [67]), calculated as: $\sqrt{\sum(x_i - M)^2/(n - 1)}$, where $M$ represents the mean target load achieved over a series of trials and $x_i$ represents the load achieved on a single trial.

Both accuracy and precision are important in quantifying how effective users are in achieving a trained target load. For example, if a user has higher accuracy but lower precision (i.e., high std. dev. of load error), this would indicate a lack of overall control.
However, if users have low accuracy and high precision, this might indicate a scheme systematically trains under or over loading that could be compensated for using a simple offset. Signed accuracy is also important in identifying training methods with biases, which would manifest in systematically positive or negative loading errors across the population.

Finally, the NASA TLX questionnaire was administered at the end of each session to assess the perceived workload of each feedback scheme. The test was administered on 7-point scale (0=min., 6=max.).

3.4.6 Analysis

Normal objective data was analyzed using a two-factor repeated-measure ANOVA, with feedback type (7 levels) and target load weight (4 levels) as within-subject factors; post-hoc tests used the Holm-Bonferroni procedure. Ordinal data (i.e., NASA TLX questionnaires), and objective data failing the Shapiro-Wilk test for normality (i.e. data that were not normally distributed), were analyzed using the Friedman test; post-hoc tests were conducted using Holm-Bonferroni corrected Wilcoxon Signed-Rank tests. In all cases, $\alpha = .05$, unless a test adjusted it for multiple comparisons. Interaction effects were tested when parametric tests could be used on both independent variables.

3.5 Results

Below, the presentation of the results is organized around each phase of the experiment (training or recall), and then subjective analysis is provided for each feedback scheme.
To contextualize these results, recall that a 'best' feedback scheme would be both accurate (have low absolute mean load error) and precise (have low standard deviation of load error). Accuracy and precision were considered both during training and recall to provide a complete understanding of how well a feedback scheme regulates the task (i.e., performance while using the feedback) and how well it trains an internal model (i.e., performance after feedback has been removed). Various target loads were included to make sure that the feedback schemes perform uniformly under different loading conditions. Since the impact of target load on compliance is not the primary focus of this work, these results are presented in the body of the thesis but accompanying figures have been moved to Appendix A: Plots of Loading Error by Target Load.

3.5.1 Training Phase

Accuracy. Absolute mean load error by feedback violated the normality assumption. There was an effect of feedback scheme on absolute mean load error ($H_6 = 40.6, p < 0.00001$); see Figure 3.3A. Post hoc tests revealed that Scale training featured significantly lower valued errors than all feedback schemes other than Visual Full, while Visual Full outperformed all but Scale and Audio Full. There was an effect of target load on absolute mean load error ($F_{3,39} = 3.04, p = 0.04$). However, post hoc tests did not reveal any difference between the levels of target load.

Signed Accuracy: Signed mean load error by feedback violated the normality assumption. There was an effect of feedback scheme on signed mean load error ($H_6 = 33.9, p <$
0.00001); see Figure 3.4A. Scale training was significantly different than all but Audio Bandwidth and Visual Terminal, Visual Terminal was also significantly different than Audio Full and Haptic. There was no effect of target load on signed mean load error \((F_{3,39} = 0.2, p = 0.9)\).

**Precision.** Std. dev. of load error by feedback violated the normality assumption. There was an effect of feedback scheme on std. dev. of load error \((H_6 = 72.3, p < 0.00001)\); see Figure 3.5A. Post hoc tests revealed significant differences in std. dev. of load error between every pairing of feedback scheme except between Scale and Visual Full, and between the three Bandwidth schemes. There was no effect of target load on std. dev. of load error \((F_{3,39} = 1.54, p = 0.22)\).

### 3.5.2 Recall

**Accuracy.** Absolute mean load error by feedback violated the normality assumption. There was no effect of feedback scheme on absolute mean load error \((H_3 = 12.5, p = 0.052)\); see Figure 3.3B. There was no effect of target load on absolute mean load error \((F_{3,39} = 0.726, p = 0.54)\).

**Signed Accuracy.** There was a main effect of feedback scheme on signed mean load error \((F_{6,78} = 4.41, p = 0.0006)\); see Figure 3.4B. Post hoc tests revealed that Visual Terminal was significantly different than Scale, Visual Bandwidth, and Haptic; Audio Full was also significantly different than Scale. There was no main effect of target load on signed mean load error \((F_{3,39} = 0.8, p = 0.5)\). The test for an interaction effect between target load and feedback scheme was not significant \((F_{18,234} = 0.88, p = 0.60)\).
**Precision.** Std. dev. of load error by feedback violated the normality assumption. There were no effects of feedback scheme ($H_3 = 9.5, p = 0.15$) or target load ($F_{3,39} = 0.36, p = 0.78$) on std. dev. of load error; see Figure 3.5B.

![Graph A](image1.png)  
**Figure 3.3. A)** Absolute mean loading error during training, aggregated by feedback scheme. **B)** Absolute mean load error during recall, aggregated by feedback scheme.
Figure 3.4. A) Signed mean loading error during training, aggregated by feedback scheme. B) Signed mean load error during recall, aggregated by feedback scheme.
3.5.3 Learning Effects

Performance was also examined over experimental sessions (7 sessions, each with 1 of the randomly presented feedback schemes), to examine whether schemes that came later in the experiment would perform better due to participants becoming better at the task (independent of feedback scheme). All three dependent variables by session number violated the normality assumption. Although there was a trend toward lower absolute mean recall error in later sessions, this trend was not significant ($H_0 = 5.8, p = 0.44$); see Figure 3.6A. Session number also did not significantly affect recall precision ($H_0 = 2.7, p$
= 0.84) or signed mean recall error ($H_6 = 4.0, p = 0.67$); see Figure 3.6B and C respectively.

Figure 3.6. A) Absolute mean loading error during recall, aggregated by session number. B) Signed mean loading error during recall, aggregated by session number. C) Standard deviation of loading error during recall, aggregated by session number.
3.5.4 Subjective Measures

TLX evaluations were used to evaluate perceptions of each feedback style in terms of workload. TLX scores were obtained by aggregating all standard TLX questionnaire items. There was an effect of feedback scheme on workload ($H_6 = 15.47, p = 0.017$); see Figure 3.7. Post hoc tests revealed no significance between groups.

![Task Load by Feedback Scheme](image)

Figure 3.7. Total TLX (task load) score ± standard error for each feedback scheme. Higher scores indicate greater workload for the feedback scheme; lower is better.

3.5.5 Summary of Results

1. Visual Terminal feedback outperformed most concurrent feedback schemes in signed accuracy during recall, suggesting that it was the most effective for learning PWB despite being the least precise and accurate during training.

2. Audio feedback schemes (given concurrently while loading on the cane) outperformed Scale (the current clinical standard) in signed recall accuracy, meaning they are better alternatives for learning PWB.
3. Scale, which represents the current clinical practice, performed among the worst of the feedback schemes for recall, suggesting that it is ineffective for learning PWB. While participants were able to regulate loads accurately during training, this did not translate to accuracy during recall.

4. Higher-resolution concurrent feedback schemes (Scale, Visual Full and Audio Full) were among the most accurate and precise for regulating loading during training. However, regulation again did not translate to recall, with the exception of Audio Full.

5. There were observed differences in precision between feedback schemes during training, but none that persisted during the recall phase.

6. PWB compliance did not vary among the different target loads (representative of clinical target loads for cane users) tested during training or during recall.

3.6 Discussion

3.6.1 Why did Visual Terminal feedback outperform concurrent feedback schemes at training the task?

Visual Terminal feedback performed the best during recall. This can be attributed to the method encouraging people to think carefully about their proprioception during each loading attempt and consider its result. When provided with concurrent feedback, participants may have focused on regulating the feedback rather than their own movements or loading. Conversely, Visual Terminal provided feedback only after attempting to hit a target load, and intuitively this lack of guidance led to poor precision.
and accuracy during training. It did, however, encourage retrospective introspection that translated to greater accuracy when recalling learned PWB loads. Similar results have been found for other motor tasks [20, 66, 74], though it is important to note that terminal feedback is not optimal for training in all situations (e.g., complex motor tasks) [85]. Our findings, therefore, provide valuable insight for the design of feedback for PWB training that can applied to clinical practice immediately. The benefits of retrospective training could be achieved by conducting the current clinical standard scale training, but hide the readings from the patient, instead having their clinician view and dictate loads after each attempt.

Since our terminal feedback was only tested with a visual representation (the numbers on the screen), it is unclear whether auditory or haptic terminal feedback could be designed to perform as well. In particular, it would likely be more difficult for users to accurately infer magnitude of their errors from intensity or frequency than it is to visually infer from numbers, making it more difficult to build an accurate internal model. However, our future work will more fully explore design alternatives for terminal feedback schemes to confirm this.

3.6.2 Why did auditory perform better than other modalities for concurrent feedback?

Both Audio Full and Audio Bandwidth performed significantly better at PWB recall than Scale, while the other concurrent feedback schemes did not. Previous analysis of neural activity during motor task learning has suggested that auditory feedback schemes divert less of the user’s attention from intrinsic (proprioceptive) feedback than visual schemes
This leads to better recall because users aren’t as dependent on the augmented feedback to perform the task.

It has been suggested that haptic feedback may actually be best for motor learning because it stimulates similar cognitive pathways to proprioception [40]. However, the multiple-resource theory suggests that haptic feedback may compete with proprioceptive feedback for cognitive resources, inhibiting the formation of an internal model [82].

The lack of clarity around these sensory systems demonstrates the importance of testing different modalities and alternatives for systems that, like ours, seek to improve the design of feedback for motor learning tasks. It is also important to understand the relative performance of different modalities when designing for users with different abilities (cognitive, visual, auditory, etc.) or different environments (quiet home vs. busy street).

### 3.6.3 Why did Scale perform poorly?

The Scale feedback scheme performed among the worst for recall, suggesting that it is particularly poor at building an internal model of the task, despite its common use in clinical practice. As a visual technique, it too may have suffered from an overemphasis on real-time regulation during the training phase.

Scale’s poor performance is particularly interesting, since Visual Terminal (the best performing scheme) was essentially the same thing but presented post-movement. The simplicity and unintuitive nature of presenting feedback post-task may be why experimenting with different feedback delays has been overlooked, and may explain why Scale, rather than Visual Terminal, is the current clinical standard.
3.6.4 Why did the performance of concurrent visual feedback during training not transfer to recall?

The guidance hypothesis indicates that frequent augmented feedback becomes part of the internal model of the motor task, degrading retention when it is removed [59, 68, 69]. It is likely that with the high bandwidth of concurrent visual feedback, guidance was more prevalent (i.e., it was easier to accurately and precisely achieve the target) and recall of the skill was inhibited.

3.6.5 Would the results translate to gait and other non-stationary situations?

By training and assessing recall on the same task (stationary PWB), we have been able to isolate the impact of three feedback characteristics on internal model formation. Previous work in the development of internal models for motor control tasks suggests that the more accurate the model, the more readily it can be transferred to other situations. Our results indicate that both terminal and concurrent auditory feedback are best for building robust internal models and should more readily transfer to other situations than current clinical practices.

3.7 Chapter Summary

In this chapter a variety of feedback alternatives were explored for training the stationary PWB skill. It was found that while high resolution concurrent feedback regulated the skill well during training, retrospective feedback was superior at training the skill for unassisted recall. This experiment considered only unimodal feedback schemes; the
following chapter will assess a multimodal training method combining the two best-performing unimodal schemes.
4 Multimodal Study

The previous chapter focused on comparing augmented feedback schemes that each targeted a single modality. In this chapter, whether or not there may be additive benefit through the combination of feedback schemes targeting different modalities when learning stationary PWB is considered. The following sections outline the motivations behind this study of multimodal feedback schemes, describe the experiment, present its results, and discuss the implications of these results.

4.1 Motivation

The results of the previous chapter suggest that terminal feedback is most appropriate for training PWB, while audio is the best modality for real-time regulation. Real-time feedback, though inferior at training the skill, vastly outperformed terminal training schemes at PWB regulation. Since they operate on different time scales, these two feedback schemes can easily be combined to create a multimodal feedback scheme. This chapter investigates whether a multimodal feedback scheme could merge the regulation and recall advantages of real-time and retrospective feedback respectively. Such a result would contradict the previous chapter’s conjecture that terminal feedback is more effective due to its promotion of proprioceptive reliance during training.

4.2 Experiment-Specific Background

4.2.1 Frequency of Feedback

Few studies have investigated motor skill retention after concurrent audio training, which may intrinsically promote kinesthetic learning rather than the guidance that concurrent
visual feedback suffers from [74]. A multimodal feedback scheme utilizing audio feedback may therefore offer real-time regulation advantages in addition to strong skill retention. However, if audio feedback diverts attention from proprioception, the proposed multimodal feedback may lose the retention benefits of terminal visual and underperform unimodal terminal training during recall.

4.2.2 Multiple Resource Theory

Processing and responding to feedback requires cognitive resources, of which humans have a limited supply. Overloading these resources may inhibit learning [90]. It has been suggested, however, that each sensory system has some amount of cognitive resources to which it has sole access [82]. This Multiple Resource Theory reveals an advantage for multimodal feedback; targeting multiple senses might allow cognitive load to be distributed across more resources, providing an additive effect to learning the new motor skill. More information could therefore be processed via a multimodal feedback scheme than from a single modality alone, and unreliable or imprecise information could be improved through other senses [74]. This may improve the accuracy of training interventions while also limiting the likelihood of guidance.

4.3 Experiment

The goal of the study was to compare the performance of the best unimodal feedback scheme to a multimodal scheme. Previous studies indicate that multimodal feedback is most effective when each modality targets a different parameter of the task [73]. Since this experimental task has a single parameter of interest, cane load, each modality was
instead chosen to target a different temporal granularity of the parameter so as to reduce cognitive load [76, 81]. This motivated a feedback scheme combining the best real-time alternative, Audio Full, with Visual Terminal, which features a different modality and provides feedback retrospectively, after each training attempt. This multimodal, mixed delay feedback scheme is referred to as Mixed.

The study consisted of two sessions, each conducted on a different day. Each session tested one feedback scheme with one target load and lasted approximately 15 minutes. Due to the lack of differences between target loads seen in the previous study and to minimize the learning effect, each session utilized a different target load (10% or 14% of the participant’s body weight), which were balanced in order of presentation to participants. Sessions occurred on different days to minimize the learning effect and participant fatigue. Sessions were not required to be on consecutive days, nor were they required to be conducted at the same time of day. Each participant received a $10 honorarium.

4.3.1 Participants

Twenty healthy university students and staff (10 identified as female and 10 as male; mean age: 23.3 years; min: 20; max: 34; s.d.: 3.4) volunteered for the experiment, which was advertised through a campus email list and posted fliers. Exclusion criteria were an inability to lean body weight on one leg or inability to support 20% of body weight through their dominant arm; no interested participants were excluded.
Demographic information such as age and dominant hand was collected at the beginning of the first session, along with body weight and height for calibrating the target loads and cane height respectively. Two participants reported being left hand dominant, with the remaining 18 being right hand dominant. Mean body weight was 151.6 lb with a minimum of 107 lb, a maximum of 197 lb, and 27.9 lb standard deviation. Ten participants reported having no prior experience using assistive devices, 6 reported less than one week of experience, 2 reported between one week and one month, and 2 reported between one month and one year.

4.3.2 Experimental Task

The experimental routine was identical to that of the previous study, which was designed to match clinical weight-bearing training sessions.

4.3.3 Procedure

The experimental procedure of this study followed that outlined in the Procedure section of the Stationary Loading Study, with one difference: there was only one target load per session (10% or 14%) with the feedback/target orderings pseudo-randomized across participants.

4.3.4 Study System

The same stationary “smart” cane system was used as in the previous study. See the Study System section of the Stationary Loading Study chapter for a full description.
4.3.5 Data

The relevant dependent variables, along with their definitions and collection criteria, were again identical to those outlined in the Data section of the Stationary Loading Study chapter.

4.3.6 Analysis

Normal objective data was analyzed using a repeated-measures ANOVA. Due to incomplete feedback-target pairings for each participant, feedback type (2 levels) and target load weight (2 levels) were analyzed separately as within-subject variables; post hoc tests used the Holm-Bonferroni procedure. Ordinal data (i.e., NASA TLX questionnaires), and objective data failing the Shapiro-Wilk test for normality, were analyzed using Wilcoxon Signed-Rank tests. In all cases, $\alpha = .05$, unless a test adjusted it for multiple comparisons.

4.4 Results

The results provided below were organized around each phase of the experiment (training or recall), and then subjective analysis is presented for each feedback scheme.

To again contextualize these results, recall that a 'best' feedback scheme would be both accurate (have low absolute mean load error) and precise (have low standard deviation of load error). Accuracy and precision were considered both during training and recall, to provide a complete understanding of how well a feedback scheme regulates the task (i.e., performance while using the feedback) and how well it trains an internal model (i.e., performance after feedback has been removed). Interaction effects were not considered
due to each participant only completing two of the four possible feedback/target combinations. Like the previous chapter, plots of results by target load can be found in Appendix A: Plots of Loading Error by Target Load.

### 4.4.1 Training Phase

**Accuracy.** Absolute mean load error, both by feedback and by target load, violated the normality assumption. There was an effect of feedback scheme on absolute mean load error \( (p = 0.0004) \); see Figure 4.1A. Visual Terminal feedback featuring significantly higher absolute mean load error during training than Mixed feedback. There was no effect of target load on mean load error \( (p = 0.39) \).

**Signed Accuracy.** Signed mean load error, both by feedback and by target load, violated the normality assumption. There was no effect of feedback scheme on signed mean load error \( (p = 0.26) \); see Figure 4.2A. There was no effect of target load on signed mean load error \( (p = 0.07) \).

**Precision.** Std. dev. of load error, both by feedback and by target load, failed the normality assumption. There was an effect of feedback scheme on std. dev. of load error \( (p < 0.00001) \), with Mixed significantly lower than Visual Terminal; see Figure 4.3A. There was no effect of target load on std. dev. of load error \( (p = 0.33) \).

### 4.4.2 Recall

**Accuracy.** Absolute mean load error, both by feedback and by target load, violated the assumption of normality. There was no effect of feedback scheme on absolute mean load error.
error \((p = 0.13)\); see Figure 4.1B. There was no effect of target load on absolute mean load error \((p = 0.9)\).

Signed Accuracy. There was no effect of feedback scheme on signed mean load error \((F_{1,19} = 1.93, p = 0.18)\); see Figure 4.2B. There was no effect of target load on signed mean load error \((F_{1,19} = 4.02, p = 0.06)\).

Precision. Std. dev. of load error, both by feedback and by target load, violated the normality assumption. There were no effects of feedback scheme \((p = 0.93)\) or target load \((p = 0.43)\) on std. dev. of load error; see Figure 4.3B.

![Figure 4.1](image1.png)

**Figure 4.1.** A) Absolute mean load error during training, aggregated by feedback scheme. B) Absolute mean load error during recall, aggregated by feedback scheme.
Figure 4.2. A) Signed mean load error during training, aggregated by feedback scheme. B) Signed mean load error during recall, aggregated by feedback scheme.
Figure 4.3. A) Standard deviation of load error during training, aggregated by feedback scheme. B) Standard deviation of load error during recall, aggregated by feedback scheme.

4.4.3 Learning Effects

Performance was also examined over experimental session (2 sessions, each with 1 randomly presented feedback scheme), to examine whether schemes that came later in the experiment would perform better due to participants becoming better at the task (independent of feedback scheme). Absolute mean load error by session during recall violated the normality assumption and featured no significant difference between the two groups ($p = 0.19$); see Figure 4.4A. There was no effect of session number on signed mean error ($F_{1,19} = 0.001$, $p = 0.974$); see Figure 4.4B. Std. dev. of load error by session
during recall violated the normality assumption and featured no significant difference between the two groups ($p = 0.39$); see Figure 4.4.

**Figure 4.4.** A) Absolute mean load error during recall aggregated by session number. B) Signed mean load error during recall aggregated by session number. C) Standard deviation of load error during recall aggregated by session number
4.4.4 Subjective Measures

The NASA TLX was again used to evaluate perceptions of each feedback style in terms of perceived load. A single TLX score (which represents task load) was obtained by aggregating all questionnaire items. There was no effect of feedback scheme on workload \((H_1 = 0.05, p = 0.82)\); see Figure 4.5.

![Task Load by Feedback Scheme](image)

Figure 4.5. Total TLX (task load) score ± standard error for each feedback scheme. Higher scores indicate greater workload for the feedback scheme. Lower is better.

4.4.5 Summary of Results

1. Despite a trend toward lower mean recall load error after Visual Terminal training, the difference was not found to be significant.

2. Multimodal, mixed timescale feedback outperformed terminal feedback for both accuracy and precision of loading regulation during training.

3. There were observed differences in precision between feedback schemes during training, but none that persisted during the recall phase.
4. PWB compliance did not vary among the different target loads (representative of clinical target loads for cane users) tested during training or during recall.

5. There was no significant learning effect between the first and second sessions, and there was no effect of conditions on perceived workload.

4.5 Discussion

4.5.1 Why was there no difference between the feedback schemes in recall load error?

Although there was a trend toward lower absolute and mean recall error with Visual Terminal training, differences between the feedback groups were not significant. At worst, one would expect Mixed feedback to perform similar to Audio Full as seen in the Stationary Loading Study. In the previous study, there were no significant differences between Visual Summary and Audio Full in terms of recall error, so this experiment increased the sample size to increase statistical power. One must recall, however, that these studies utilized a cane with torque-based load quantification. Slight variations in grip and pressure distribution throughout the hand could have changed the user’s moment arm, resulting in slight differences in load readings for the same actual load through the cane. Although participants were coached to “always hold the cane the same way,” intra- and inter-subject variability is a confounding factor. It is possible that, even with this sample size, the difference in effectiveness between these two feedback schemes is not large enough to be significant given the lack of precision in the current instrument.
4.5.2 Why did Multimodal feedback outperform Visual Terminal during training?

The multimodal scheme featured real-time audio feedback in addition to the retrospective feedback of Visual Terminal. This additional concurrent feedback allowed users to easily load to near the target with every attempt, whereas terminal feedback alone relied on “guessing and checking.” Because of the lack of immediate feedback while performing the task with Visual Terminal it intuitively performed worse during the training phase.

4.6 Chapter Summary

In this chapter a multimodal, mixed timescale PWB training method was compared to the best performing unimodal scheme from the previous chapter. The multimodal scheme, which featured both concurrent and retrospective components, was found to better regulate the skill during training. There was a trend toward better performance after unimodal retrospective feedback training during recall, although not significantly so. It is postulated that technological limitations made it difficult to draw conclusions regarding this performance. With evidence to support the idea that multimodal feedback does not seem to provide clear benefit over unimodal feedback schemes, we now turn our attention to the next major question: do better stationary PWB skills acquired by augmented feedback training improve the compliance of PWB during gait? The next chapter therefore assesses skill transfer from stationary training methods to a dynamic gait context.
5 Gait Study

In this chapter, the skill transfer of various PWB training alternatives is assessed by measuring compliance during training, stationary recall, and gait recall. The following sections outline the motivations behind this study, describe the experiment, present its results, and discuss the implications of these results along with potential future work.

5.1 Motivation

The results of the first two studies suggest that terminal feedback is superior for learning the stationary partial weight-bearing skill. Previous work indicates that stationary PWB training does not transfer well to gait recall [16, 20, 31], although no previous studies have examined how gait-based training might perform during stationary recall. This is particularly important for assessing the feasibility of gait-based training. While gait is often considered problematic in terms of the potential for leading to injury, systematically over- or under-loading a lower limb while standing still could potentially lead to similar problems [86, 96]. Previous work has not considered whether gait-based methods might transfer well, or at all, to recall in stationary context. Furthermore, terminal stationary training methods, like the one that performed best in the first study, have not been assessed for retention during gait (aside from the hand-under-foot method, which has poor precision [20]).

This chapter investigates the transfer of PWB skill between contexts (gait training to stationary recall and stationary training to gait recall) to determine whether improved learning of the stationary PWB skill leads to better compliance during gait, and how
stationary training compares to gait training for PWB recall in both contexts. This study addressed important gaps in understanding about which methods should be pursued by technology designers and by clinicians.

5.2 Experiment-Specific Background

5.2.1 Skill Transfer

Humans perform a variety of motor tasks every day. The same hand orientation and grip are used when holding a cup, for example, regardless of whether it is full or empty. Each situation (from the weight of the mug), however, leads to different torques that must be countered by varying levels of supination (i.e., how much the wrist rotates). There are infinite specific scenarios of motor tasks, so humans can’t feasibly learn a task under every possible circumstance (e.g., every minute variation of cup weight). We can, however, learn a set of core tasks and transfer or adapt these skills to similar contexts. A robust internal model is critical for skill transfer as it allows one to predict the impact of various factors on the task and adequately account for them in their execution strategy [50]. In general, more variation during training allows one to learn the cause-and-effect relationship of various parameters on their performance, leading to better skill transfer [9, 39]. Though practice was not intentionally varied in this experiment, concurrent feedback essentially allows users to explore the entire loading space during each attempt. This contrasts with retrospective feedback, which only informs users of a discrete load value aggregated over the latter portion of their attempt. Though the Visual Terminal feedback
scheme outperformed Scale in the Stationary Loading Study, Scale’s variation during practice may offer an advantage when transferring the skill to gait.

5.2.2 **Steady-State Gait**

As described in the Mechanics of Gait section in the Background chapter, gait is a periodic means of locomotion. People don’t typically walk continuously all day however; there is a great deal of starting and stopping when gait is performed in everyday activities. Previous work has indicated that people reach an equilibrium state with constant cadence within the first step of regular gait [6], but the same is not true for assisted gait. Hustedt et al. found that it took 40 to 50 strides for crutch users to acclimate and reach a steady state of partial weight-bearing [31]. No evidence could be found to suggest that this effect would be any different when using canes as an assistive device.

5.3 **Experiment**

The goal of this study was to compare the effectiveness of stationary and gait PWB training methods at preparing cane users to recall the skill both while stationary and while walking. Previous studies indicate that gait feedback is more effective than stationary methods when recall is assessed during gait [20, 30, 31], but it isn’t clear whether this is due to inferior stationary training methods or poor transfer of the skill to the gait context. This motivated us to compare the best performing stationary scheme found in the previous studies (Visual Terminal) and the current clinical standard stationary scheme (Scale) with a gait-based scheme that emulates Visual Terminal, called Gait Summary.
This Gait Summary training method features a bar plot illustrating the peak load of each stride after a series of strides.

The study consisted of three sessions on different days. Each session tested one feedback scheme for one target load, lasting approximately 15 minutes. Due to the lack of differences between target loads seen in the previous studies and to minimize the learning effect, each session utilized a different target load (10%, 12%, or 14% of the participant’s body weight). Sessions occurred on different days to minimize the learning effect and participant fatigue. Sessions were not required to be on consecutive days, nor were they required to be conducted at the same time of day. Each participant received a $10 honorarium.

5.3.1 Feedback Schemes

The best and worst performing feedback styles from the first study, Visual Terminal and Scale, were emulated in a mobile application for this study; see Figure 5.1.
Figure 5.1. Scale and Visual Terminal feedback screen on mobile device. For Scale feedback, the participant was shown the screen as the load number updated in real time. For Visual Terminal feedback, participants were shown a mean load after their attempt. The red box in the middle of the screen was a button used by the experimenter to denote completion of a loading attempt.

A gait training method was developed to offer retrospective batched feedback. Participants took several strides with a wireless smart cane (described in the Study System section below), after which they viewed a bar chart illustrating the peak load of each stride; see Figure 5.2. This facilitated terminal feedback without disrupting the flow of gait after each stride, along with comparisons between the stride loads.
Figure 5.2. Gait feedback screen on mobile device. Each bar indicates the peak load through the cane for a given attempt, while the green line indicates the target load. Participants were shown the screen after each pass of the 12 ft course.

5.3.2 Participants

Eighteen healthy university students and staff (6 identified as female and 12 as male; mean age: 23 years; min: 18; max: 35; s.d.: 2.9) volunteered for the experiment, which was advertised through a campus email list and posted fliers. Exclusion criteria were an inability to lean body weight on one leg or inability to support 20% of body weight through dominant arm; no interested participants were excluded.
Demographic information such as age and dominant hand was collected at the beginning of the first session, along with body weight and height for calibrating the target loads and cane height respectively. One participant reported being left hand dominant, with the remaining 17 being right hand dominant. Mean body weight was 158.9 lb with a minimum of 112 lb, a maximum of 203 lb, and 27.1 lb standard deviation. Six participants reported having no prior experience using assistive devices, 9 reported less than one week of experience, 1 reported between one week and one month, and 2 reported between one month and one year.

5.3.3 Experimental Task

The experimental routine was similar to that of the previous studies, comprising of the same phases (train, distract, and recall). The primary task of each session was loading an instrumented cane with the dominant arm. Participants would train with the session’s feedback scheme to load the cane to a defined target load. Depending on the feedback scheme, this was either while stationary or while walking. They were then distracted to clear their short-term memory, and subsequently asked to recall the target load without assistance. Regardless of training method, recall was assessed both while stationary and while walking.

5.3.4 Procedure

During the first session, participants were explained the experimental procedure, provided informed consent, and completed a demographic questionnaire. They were also weighed on an analog bathroom scale to allow proper cane calibration.
Each session featured a different feedback scheme and target load. The cane length was adjusted to properly fit the participant at the beginning of each session according to best clinical practice. The session began with a training phase: the participant would lean on the cane while receiving feedback assistance to help them achieve one of three target loads (10%, 12%, or 14% of their body weight) and be informed that they would later need to recall this load without assistance.

Gait training consisted of 4 passes of a 12-ft walkway. Each pass featured approximately 3 strides which were visualized on the screen of a paired phone. The peak load of each stride was displayed, with the target load superimposed across all strides; see Figure 5.2. The plot was cleared between passes, after the participant had a chance to view and comprehend the information.

During stationary training (Scale and Visual Terminal), participants made 10 stationary loading attempts with feedback assistance. This number was selected to offer equivalent training opportunities to that of Gait training, which featured approximately 12 attempts and offered feedback 4 times. They verbally notified the experimenter when the target was achieved and raised the cane off the ground between attempts. After 5 attempts, the cane was handed to the experimenter while the participant rubbed their hands together and counted backward from 30. This cleared their haptic memory [19] and partitioned training to make it comparable to the gait-based scheme.

After training, participants were distracted by counting backwards from 50 while rubbing their hands together. This distraction phase was followed by stationary recall: 10 attempts at loading the cane to the target without feedback assistance. Like with stationary
training, participants counted backward from 30 while rubbing their hands after 5 attempts.

After stationary recall, participants were once again distracted by counting backwards from 50 while rubbing their hands together. This distraction phase was followed by gait recall: walking with the cane while attempting to match the peak load of each stride to the target load. Participants walked around a hallway loop which yielded approximately 140 strides.

Sessions concluded with a TLX questionnaire and experimenter-solicited comments regarding the session and/or comparing the three schemes. The presentation of target load and feedback orderings were balanced across the 14 participants.

5.3.5 Study System

In order to evaluate our feedback schemes, we used a wireless “smart cane” system, developed at the Health Technologies Laboratory at UNB, as described in [18]. This device streamed load data via Bluetooth Low Energy (BLE) to a customized Android application for data logging and feedback rendering. Using a calibration curve, the ADC load value was mapped to a load in pounds.

A sampling rate of 100 Hz allowed high resolution loading data to be reliably collected by the application. Each load reading was logged to a comma separated variable file along with the timestamp, target load, anonymized participant ID, feedback condition, and experimental phase (training, stationary recall, or gait recall). If the participant was undergoing a training phase, load was also used to render feedback. For gait training,
strides were segmented by non-zero load series lasting at least 0.4 s with at least 0.4 s of zero load between them. A “zero threshold” of 3 lb was set to mitigate the impact of noise on this algorithm.

5.3.6 Data

The relevant dependent variables, along with their definitions, are identical to those outlined in the Data section of the Stationary Loading Study chapter. The collection criteria for raw data differs, however, due to the inclusion of gait training and recall.

Load error for the stationary task was quantified as the difference between the target load and the mean of load samples for 1 second (i.e., 100 samples) after the participant indicated that they had reached the target load. Each stationary phase (either stationary training or stationary recall) consisted of 10 attempts to match the target load.

Load error for the gait task was quantified as the difference between the target load and the peak load of a stride. Gait training consisted of approximately 12 strides, with the exact number depending on the number of strides a participant took to cross the 12 ft platform (mean: 12.5 strides; min: 8; max: 16; std. dev.: 1.9). Gait recall consisted of approximately 140 strides, with the exact number depending on the number of strides a participant took to complete the course (mean: 142 strides; min: 69; max: 205; std. dev.: 29.3).

5.3.7 Analysis

Normal objective data was analyzed using a repeated-measure ANOVA. Due to incomplete feedback-target pairings for each participant, feedback type (3 levels) and
target load weight (3 levels) were analyzed separately as within-subject variables; post hoc tests used the Holm-Bonferroni procedure. Ordinal data (i.e., NASA TLX questionnaires), and objective data failing the Shapiro-Wilk test for normality, was analyzed using Friedman test; post hoc tests were conducted using Wilcoxon Signed-Rank tests. In all cases, $\alpha = .05$, unless a test adjusted it for multiple comparisons.

5.4 Results

The following results are organized around each phase of the experiment (training, stationary recall, and gait recall), and then subjective analyses are presented for each feedback scheme.

To again contextualize these results, recall that a 'best' feedback scheme would be both accurate (have low mean load error) and precise (have low standard deviation of load error). Accuracy and precision were considered both during training and recall to provide a complete understanding of how well a feedback scheme regulates the task (i.e., performance while using the feedback) and how well it trains an internal model (i.e., performance after feedback has been removed). By considering recall in both gait and stationary tasks, how well PWB skill transfers to other contexts (e.g., stationary training to gait recall), an indicator of more robust internal models, can be evaluated. Interaction effects were not considered due to each participant only completing three of the nine possible feedback/target combinations. Unless otherwise stated, continuous variable distributions were normal. As in previous chapters, plots of results by target load can be found in Appendix A: Plots of Loading Error by Target Load.
5.4.1 Training Phase

Accuracy. Absolute mean load error, both by feedback and by target load, violated the normality assumption. There was an effect of feedback scheme on absolute mean load error ($H_2 = 24.1, p < 0.00001$); see Figure 5.4A. Scale and Visual Terminal feedback featured significantly lower mean load error during training than Gait Summary feedback. There was no effect of target load on absolute mean load error ($H_2 = 0.33, p = 0.85$).

Signed Accuracy. Signed mean load error, both by feedback and by target load, violated the normality assumption. There was an effect of feedback scheme on signed mean load error ($H_2 = 8.8, p = 0.01$); see Figure 5.5A. Scale and Visual Terminal feedback featured significantly lower signed mean load error during training than Gait Summary feedback. There was no effect of target load on signed mean load error ($H_2 = 0.11, p = 0.95$).

Precision. Std. dev. of load error, both by feedback and by target load, violated the normality assumption. There was an effect of feedback scheme on std. dev. of load error ($H_2 = 32.4, p < 0.00001$); see Figure 5.6A. Post hoc tests revealed significantly lower std. dev. of load error while training with Scale than the other methods, with Visual Terminal also having a lower value than Gait Summary. There was no effect of target load on std. dev. of load error ($H_2 = 0.11, p = 0.95$).
5.4.2 Stationary Recall

Accuracy. Absolute mean load error, both by feedback and by target load, violated the normality assumption. There was no effect of feedback scheme on absolute mean load error ($H_2 = 8.44, p = 0.01$); see Figure 5.4B. Visual Terminal training resulted in significantly lower absolute mean load errors during stationary recall than both Scale and Gait Summary training. There was no effect of target load on absolute mean load error ($H_2 = 5.44, p = 0.07$).

Signed Accuracy. Signed mean load error, both by feedback and by target load, violated the normality assumption. There was an effect of feedback scheme on signed mean load error ($H_2 = 13.8, p = 0.001$); see Figure 5.5B. Every feedback pairing differed significantly. There was no effect of target load on signed mean load error ($H_2 = 4.3, p = 0.11$).

Precision. Std. dev. of load error, both by feedback and by target load, violated the normality assumption. There were no effects of feedback scheme ($H_2 = 2.3, p = 0.3$) or target load ($H_2 = 4.8, p = 0.09$) on std. dev. of load error; see Figure 5.6B.

5.4.3 Pre-Steady-State Gait Recall

Accuracy. Absolute mean load error by feedback scheme violated the normality assumption. There was an effect of feedback scheme on absolute mean load error ($H_2 = 14.3, p = 0.0008$); see Figure 5.3A. Gait Summary had significantly lower absolute mean load error than both Scale and Visual Terminal.
Signed Accuracy. There was an effect of feedback scheme on signed mean load error \((F_{2,34} = 23.1, p < 0.000001)\); see Figure 5.3B. Gait Summary had a lower signed mean load error than both Scale and Visual Terminal.

Precision. Std. dev. of load error by feedback scheme violated the normality assumption. There was no effect of feedback scheme on std. dev. of load error \((H_2 = 4, p = 0.14)\); see Figure 5.3C.
Figure 5.3. A) Absolute mean load error during the first 50 strides of gait recall, aggregated by feedback scheme. B) Signed mean load error during the first 50 strides of gait recall, aggregated by feedback scheme. C) Standard deviation of load error during the first 50 strides of gait recall, aggregated by feedback scheme.
5.4.4 Steady-State Gait Recall

Accuracy. Absolute mean load error, both by feedback and by target load, violated the normality assumption. There was no effect of feedback scheme on absolute mean load error ($H_2 = 0, p = 1$); see Figure 5.4C. There was no effect of target load on absolute mean load error ($H_2 = 2.33, p = 0.31$).

Signed Accuracy. There was an effect of feedback scheme on signed mean load error ($F_{2,34} = 15.8, p < 0.0001$); see Figure 5.5C. Gait Summary had significantly different signed mean load error than Scale and Visual Terminal. There was no effect of target load on signed mean load error ($F_{2,34} = 0.49, p = 0.62$).

Precision. There was an effect of feedback scheme on std. dev. of load error ($F_{2,34} = 2.03, p = 0.15$); see Figure 5.6C. Gait Summary featured a lower std. dev. of load error than Scale. There was no effect of feedback scheme on std. dev. of load error ($F_{2,34} = 2.59, p = 0.09$).
Figure 5.4. A) Absolute mean load error during training, aggregated by feedback scheme. B) Absolute mean load error during stationary recall, aggregated by feedback scheme. C) Absolute mean load error during steady-state gait recall, aggregated by feedback scheme.
Figure 5.5. A) Signed mean load error during training, aggregated by feedback scheme. B) Signed mean load error during stationary recall, aggregated by feedback scheme. C) Signed mean load error during steady-state gait recall, aggregated by feedback scheme.
Figure 5.6. A) Standard deviation of load error during training, aggregated by feedback scheme. B) Standard deviation of load error during stationary recall, aggregated by feedback scheme. C) Standard deviation of load error during steady-state gait recall, aggregated by feedback scheme.
5.4.5 Learning Effects

Performance was again evaluated over experimental session (3 sessions, each with 1 randomly presented feedback scheme) to examine whether schemes that came later in the experiment would perform better due to participants becoming better at the task (independent of feedback scheme). Signed mean load error by session number violated the normality assumption. There was no effect of session number on signed mean load error during stationary ($H_2 = 0.11, p = 0.95$) or gait ($F_{2,17} = 0.67, p = 0.52$) recall; see Figure 5.7C and D. Absolute mean load error by session violated the normality assumption during both phases and did not feature effects during stationary ($H_2 = 2.1, p = 0.35$) or gait ($H_2 = 0.78, p = 0.68$) recall; see Figure 5.7A and B. Std. dev. of load error by session violated the normality assumption during stationary recall and did not feature an effect ($H_2 = 0, p = 1$); see Figure 5.7E. Std. dev. of load error by session did not feature an effect during the gait phase ($F_2 = 2.2, p = 0.13$); see Figure 5.7F.
5.4.6 Subjective Measures

TLX evaluations were used to evaluate perceptions of each feedback style in terms of workload. TLX scores were obtained by aggregating all standard TLX questionnaire items. There was no effect of feedback scheme on workload ($H_0 = 3.8, p = 0.15$); see Figure 5.8.
Figure 5.8. Total TLX (task loading) score ± standard error for each feedback scheme. Higher scores indicate greater workload for the feedback scheme. Lower is better.

5.4.7 **Summary of Results**

1. Scale and Visual Terminal outperformed Gait Summary feedback during training on all three measures, while Scale also outperformed Visual Terminal on precision.
2. Visual Terminal was more accurate (both signed and absolute) than both other methods during stationary recall.
3. There were no precision differences between the feedback schemes during stationary recall.
4. PWB compliance did not vary among the different target loads (representative of clinical target loads for cane users) tested during training, stationary recall, or gait recall.
5. Gait training led to underloading during both stationary and gait recall, whereas stationary training led to overloading.
6. Gait Summary training outperformed the other two schemes in accuracy at the start of gait recall, but there were no significant differences in absolute mean error during steady-state gait.

7. The most precise feedback scheme during training (Scale) yielded the least precision during gait recall, while the least precise scheme during training (Gait Summary) yielded the most precision during gait recall.

8. Though there was a trend toward better performance in later sessions, there were no significant learning effects for any dependent variable.

9. There were no significant differences in perceived task load between the feedback schemes.

5.5 Discussion

5.5.1 Why did stationary training methods outperform the gait training method during training?

Scale and Visual Terminal schemes feature more frequent feedback to the user than Gait Summary training. Frequent feedback allows users to adjust their approach and hone in on a target more quickly. Additionally, Gait Summary training errors were determined by peak stride load, which is more variable than the stationary errors which were aggregated over 1 second.
5.5.2 Why was Visual Terminal more accurate than the other methods during stationary recall?

Visual Terminal’s superiority over Scale in this phase replicated the results discussed in the Stationary Loading Study chapter. Gait Summary training required skill transfer to partial weight-bearing in a stationary context, which introduces an additional source of error to the task.

5.5.3 Why did Gait Summary training lead to underloading during recall?

Several participants noted that target loads felt lighter during gait. There are a few potential explanations for this. First, the periodic nature of gait only requires cane loads to be sustained for tenths of a second, whereas stationary loading required participants to sustain the target load for 3 seconds. The total energetic expenditure for each attempt is therefore higher while stationary than during gait. Second, vertical forces on the body oscillate through gait, with a peak at 120% BW. Any cane target load comprises a smaller percentage of this load than the maximum vertical force while stationary (i.e., 100% BW), thus may be perceived as being lighter in comparison. Finally, peak cane load occurs around midstance while only one lower limb is being used for support [18]. Even with assistance from the cane, the affected limb bears a much larger load than during stationary loading, where the burden is shared by two lower limbs. Since decreased lower limb load is assumed to be due to increased cane load, gait training may cause users to underload the cane during stationary recall because they have developed an internal model that expects higher lower limb loads. A similar effect could explain why users overloaded during gait recall after stationary training. One should note, however, that this
study considered load through the cane whereas many previous studies have found that scale training promoted overloading the lower limb [16, 20, 31]. Area of focus is known to affect performance in motor tasks [88, 91], so it is possible that stationary training biased users to overload whichever limb they focused on loading (i.e., the arm in this experiment, the leg in previous work). If a consistent loading bias was identified, clinical training methods could be modified to target a load that is higher or lower than users are expected to recall.

5.5.4 Why did Gait Summary training outperform Scale and Visual Terminal in accuracy at the start of gait recall, but not during steady-state gait?

It seems logical that Gait training would outperform the stationary methods during gait recall; there is no skill transfer necessary. While this held true for the first 50 strides, Gait Summary’s performance slipped to being comparable to the stationary methods during steady-state (>50 strides) gait recall. This can most likely be attributed to the briefness of training preventing users from developing a deeply ingrained “muscle memory” internal model of the task [54]. Their transient performance depended on an internal model in their short-term memory but faded over time as they settled into a comfortable gait pattern [31]. It is also possible that fatigue played a role in equalizing performance during steady-state gait, causing the cane load to drop in all conditions as perceived exertion increases [77]. This would cause Gait Summary’s signed mean errors to be negative (increasing absolute mean errors) while stationary loading errors, which were quite high and positive during the pre-equilibrium phase, dropped to an absolute level near that of Gait Summary.
5.5.5 Why was the most precise feedback scheme during training (Scale) the least precise during gait recall? Why was the least precise scheme during training (Gait Summary) the most precise during gait recall?

Effective PWB training requires a balance between the user experiencing the target load but also working to get there. Concurrent feedback schemes, such as Scale, allow the user to achieve the target with great precision early in training, but heavily guide them to this target. Guidance limits internal model development, depressing performance during skill recall and transfer [68]. Scale’s poor precision during gait recall indicates that it features too much guidance to develop a robust internal model of the task. The Gait Summary feedback scheme, on the other hand, forces users to rely on their proprioception while exploring the cane loading space. Furthermore, they are both training and recalling during gait, removing the need for skill transfer. Gait training therefore should be expected to build a robust internal model and avoid random errors from skill transfer, improving precision during gait recall.

5.6 Chapter Summary

In this chapter the performance of stationary and gait-based PWB training methods were assessed during training, stationary recall, and gait recall. It was found that concurrent real-time feedback regulated the skill best during training, stationary retrospective feedback trained the skill best for stationary recall, and gait-based feedback trained the skill best for gait recall.
6 General Discussion

In this chapter, commonalities and differences in the findings of all three studies are discussed, as well as the work in this thesis more generally.

6.1 Common Findings

6.1.1 Why are there precision differences during training but not stationary recall?

There seems to be a direct connection between the amount of information available and the amount of precision achieved. During training, the differences observed for precision were almost a perfect mirror of those observed for accuracy. However, unlike accuracy, the precision of all feedback schemes was roughly equal during recall. We believe this is due to the overall difficulty of the recall task, which is performed without feedback, forcing users to rely solely on proprioception. While training methods teach users how to map their proprioception to objective values (i.e., load in lb), they don’t train users to have improved consistency when loading. Better training should give users a better idea of where the target is (better accuracy during recall), but as seen in this study, they will have the same precision (indicated by standard deviation of loading attempts) during recall regardless of feedback scheme.

6.1.2 Why do variations in target load not affect performance?

These results are consistent with previous literature, which found that young asymptomatic participants have similar loading errors regardless of target load [31]. It has also been shown that BMI has more impact on loading error than target load [30];
however, our target loads and loading errors were normalized by body mass to reduce inter-subject variability.

6.1.3 Are the results applicable to a patient population?

The participants in these studies were all healthy and without known injury. While in general research should, whenever possible, solicit the participation of target populations, we have no reason to believe that our results would not hold. In this work, we were interested in identifying the techniques that will best support building an accurate internal model. We found no literature or reason to suggest that the optimal training methods (Visual Terminal for stationary and Gait Summary for gait recall) would be less effective for a patient population; the simplicity of these feedback schemes would likely be easier for some demographics to interpret and utilize than other alternatives. Users with lower limb pain may be wary of overloading their limb, but this would be independent of training scheme and the additional nociceptive (i.e. pain) intrinsic feedback may strengthen internal model load mappings. Of course, sensory deficits such as blindness necessitate the consideration of training schemes that target other senses, such as Audio Full. Our ongoing work seeks to include our target populations and seek feedback from clinicians and other stakeholders.

6.2 Implications for the Design of Feedback

These findings can be used to directly inform current clinical practice, and the design of improved technologies for training PWB. Current training practices should employ terminal (post-movement) feedback rather than concurrent (real-time) feedback for
training PWB. While current practices, such as using a scale, would still be useful, clinicians might prevent patients from viewing the scale’s readout and simply report the results of a loading attempt after it is complete.

While high-precision visual methods can assist in the real-time regulation of PWB, their practicality is questionable in a clinical use case where the focus is most likely on training internal models to apply the skill in daily life. We found in the Gait Study, however, that the benefits of training can fade over time. This contradicts previous work which found that training can last up to 24 hours [4, 16, 32], but may be explained by quicker onset of fatigue while using a cane as compared to axillary crutches. Further research is required to determine optimal training routines for long-term retention of PWB performance. Previous research suggests that randomizing target loads during practice may help users develop a more robust internal model of the task space [89]; however, care would need to be taken to ensure that target loads are feasible and realistic for the assistive device (e.g., less than 25% BW for a cane) and avoid damaging the affected lower limb [94]. Unless future work finds a training schedule for developing a more permanent internal model of the PWB task, users would have to retrain every dozen or so minutes. This could lead to frustration and abandonment of the training program, so a long-term regulation approach may be preferable if designed to be unobtrusive. This could be inconspicuous real-time feedback (e.g., vibration of a wrist-worn fitness tracker) or occasional summary feedback (e.g., occasionally viewing a plot on a connected smartphone ).
The results found here suggest that PWB regulation should consider a feedback technique similar to our Audio Full scheme since it provides similar performance during training to visual methods but has the important benefit of providing learning that performs better during recall. Furthermore, it doesn’t inhibit the user’s vision, a sense that most people rely heavily on for performing acts of daily living. Haptic feedback is also an option for discrete training, the levels and presentation of which could likely be optimized to yield regulation performance similar to that of audio feedback. Previous literature has shown that both modalities can be utilized for effective PWB regulation during gait [14, 16, 49].

Regardless of feedback scheme, PWB compliance is determined by the load through an affected lower limb. This can be quantified directly by instrumenting the foot of the affected limb, or by proxy using the load through an assistive device. Simply subtracting device load from body weight doesn’t suffice because PWB is complicated during gait by the body’s natural vertical oscillation varying vertical forces throughout the gait cycle [33]. From a usability standpoint, however, it is less intrusive and cumbersome to instrument a cane or crutch than to have users wear a pressure-sensitive boot or insole. Our target population uses these assistive devices anyway, so augmenting them to provide added value avoids confounding or inhibiting their gait or behaviors. Further research is required, however, to determine a reliable mapping between device load and limb load throughout the gait cycle.
6.2.1 **What can the results tell us about training internal models for motor control more generally?**

The results showing that terminal and auditory feedback performed best for recall indicates that PWB is a simple motor task (see the Guidance section of the Introduction chapter). Previous work has identified that simple tasks are trained best with terminal and auditory feedback, while complex tasks are not [74]. This is further supported by the poor performance of the higher resolution visual feedback (Scale) at recall. Future work could apply our feedback schemes to known simple motor tasks to confirm the schemes’ performances, and the generalizability of the simple task rule-of-thumb for feedback design.
7 Conclusions and Future Work

7.1 Stationary Loading Study

In the first experiment, we provided an initial investigation of feedback types for stationary partial weight-bearing using a smart cane. Based on the results of our study of 14 participants, terminal feedback (presented after the training action was attempted) promotes better learning of the skill than concurrent feedback (feedback offered while the skill is being attempted). Further, auditory feedback may be a more effective concurrent feedback modality than others tested. These findings can be leveraged to create a better clinical training method for this skill by simply combining aspects of the two current methods (Audio Full and Visual Terminal, described above). Rather than allowing users to view scale readings, clinicians could dictate loads to them after each attempt to reap the training benefits of terminal feedback.

By closely approximating clinical best practices and basing our designs on a solid understanding of motor control training, our findings provide new insight into how internal models can be better trained. These findings are important since prevailing wisdom on motor control training suggests improved internal models should improve compliance in a wide variety of tasks. We plan to build on these findings and use them to inform the design of both augmented feedback training and continuous intervention for partial weight-bearing. We are also interested in incorporating sensors such as EMG cuffs and pressure-mapping shoe insoles to enrich the loading information available to partial weight-bearing systems. Improved research in these directions could improve
rehabilitation outcomes and allow assistive device users the freedom to more safely complete a wide range of activities of daily living, such as walking and climbing stairs, among others.

7.2 Multimodal Study

It is difficult to draw meaningful conclusions from the Multimodal Study due to technical limitations. Like Audio Full in the Stationary Loading Study, Mixed feedback outperformed Visual Terminal during training. This is unsurprising as concurrent audio feedback facilitates regulation of the skill within each trial, as opposed to terminal only aiding users between trials. There was no clearly superior feedback style during recall however, which in itself could be a meaningful result. Mixed feedback outperformed Visual Terminal at PWB regulation and was comparable at training the skill, therefore it should be considered a more favourable PWB training scheme. Unfortunately, as described in the Study System section of the Stationary Loading Study chapter, the input device is not very reliable. Although this uncertainty effected all three studies, differences in learning were strong enough to overcome the noise in all but the Multimodal Study. There was a trend toward Visual Terminal outperforming Mixed at training the PWB skill, but it is not clear whether the lack of distinguishability between the schemes is due to them actually performing similarly or the device simply providing too noisy a signal to yield meaningful results. Consequently, firm recommendations cannot be made based on the results of this experiment.

Despite this lack of deterministic result, this line of research sheds light on the merit of investigating multimodal feedback alternatives for motor learning. Even if multimodal
approaches are found to be less than or equally effective as unimodal strategies, such studies could help identify mechanisms of learning from different modalities and/or time delays. Feedback can be designed in innumerable different ways and applied to a vast number of different tasks; there is immense value in understanding how and why certain parameters impact learning and regulation. Future work could revisit this experiment with more precise input devices and a larger sample size to more definitively determine whether the feedback schemes are comparable at training the PWB skill. It would also be interesting to include unimodal Audio Full in the study so Mixed feedback can be compared to both its components (Audio Full and Visual Terminal).

7.3 Gait Study

In the Gait Study, we examined the skill retention and transfer performance of stationary- and gait-based PWB training methods. We replicated the results of the Stationary Loading Study with retrospective stationary training outperforming concurrent stationary training at skill recall (i.e., unassisted execution of the skill in the same context in which training occurred). We also replicated that experiment’s training precision results: less frequent feedback intuitively led to lower precision during training. This finding was extended by the inclusion of Gait Summary training, which featured both the least frequent feedback (every 3-4 attempts compared to Visual Terminal’s every-attempt and Scale’s within-attempt feedback) and the least precision during training.

Beyond simply replicating previous results, this study revealed two novel findings. It was observed that Gait Summary training caused users to decrease their cane loading in both the stationary and steady-state gait recall contexts below that of the target load. It was
also observed that Gait Summary training yielded near perfect PWB accuracy during the first 50 strides of gait recall, but this compliance dropped off during steady-state gait. There is limited previous work regarding post-biofeedback PWB during gait to compare with, but it has been shown that crutch users can accurately recall target loads up to 24 hours after biofeedback training [31]. That study limited trials to 50 strides, however, and axillary crutches are easier to sustain heavy loads through [93]. It is likely that the Gait Summary training scheme developed a robust internal model of the task, resulting in impressive performance during the first portion of gait recall, but muscle fatigue or adaptation caused users’ perceived exertion to exceed their actual loading pattern [77]. This could be attributed to a lack of training under fatigued conditions preventing participants from understanding the relationship between fatigue and proprioception while loading.

In future work, more varied and extensive training could be used to build more robust internal models of the task, potentially allowing users to better adapt to fatigue [9, 39]. Furthermore, concurrent gait training methods should be compared to the summary method used in this study to determine whether terminal feedback is superior at PWB training in the gait context as the Stationary Loading Study found it to be in the stationary context. It would also be advantageous to assess lower limb loads during gait recall because that is the clinical measure of interest. Cane load simply acts as a proxy, but is not a perfect representation of the lower limb load due to varying vertical forces on the body during gait [33].
7.4 Overview

This thesis has examined feedback alternatives for training the PWB skill in both the stationary and gait contexts. The results of three experiments identified a simple method of improving current clinical training and highlighted the importance of task-specific training for motor tasks like PWB. They indicate that single-session PWB training may be infeasible, suggesting future work should focus on optimizing training schedules along with discrete continuous biofeedback regulation systems.
Bibliography


101


[80] University of Wisconsin Sports Medicine 2017. *Rehabilitation Guidelines Following Osteochondral Allograft or Autograft Transplantation (OATS).*


Appendix A: Plots of Loading Error by Target Load

Figure A.1. Stationary Loading Experiment results by Target Load.
Figure A.2. Multimodal Experiment results by Target Load.
Figure A.3. Gait Experiment results by Target Load.
Curriculum Vitae

Candidate’s full name: Ian Christopher James Smith

Universities attended (with dates and degrees obtained):

Bachelor of Science with Honours in Applied Physics (BSc) – May 2017

University of New Brunswick

Fredericton, New Brunswick, Canada

Bachelor of Computer Science (BCS) – May 2017

University of New Brunswick

Fredericton, New Brunswick, Canada

Publications: None

Conference Presentations:

 April 2019 - 16\textsuperscript{th} Annual UNB FCS Research Expo

What Moves You? Augmented Biofeedback for Enhancing Motor Learning

October 2017 - Mathematics, Statistics and Computer Science Conference (Science Atlantic)

Improving Gait Through Sensored Cane Feedback