An evaluation of daily activity recognition using on-body inertial sensors

Author:
Abdulla, Umran

Publication Date:
2015

DOI:
https://doi.org/10.26190/unsworks/18346

An evaluation of daily activity recognition using on-body inertial sensors

Umran Azziz Abdulla

A thesis submitted in fulfilment

of the requirements of the degree of

Doctor of Philosophy

School of Engineering and Information Technology

UNSW Canberra at ADFA

University of New South Wales

February, 2015

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Abdulla

First name: Umran | Other name/s: Azziz

Abbreviation for degree as given in the University calendar: 1885

School: Engineering and Information Techology | Faculty: UNSW Canberra at ADFA

Title: An evaluation of daily activity recognition using on-body inertial sensors

Abstract 350 words maximum: (PLEASE TYPE)

Considerable research has been performed exploring Activity Recognition (AR) using wearable sensor nodes such as smart phones that incorporate accelerometers, gyroscopes and magnetometers.

This thesis presents an empirical evaluation of AR using on-body sensors. It studies the recognition of 22 Activities of Daily Living using either wearable sensors that have fixed locations on the subject's body or smart phones carried by the subject. The differences between the two sensor configurations are explored as well as parameters for computationally efficient and accurate AR.

Initially, fundamental classifier settings that impact AR accuracy are explored. These include evaluating the performance of acceleration, rotational velocity and orientation derived features (the three are referred to as sources). In addition, minimum sampling frequencies, window size, window overlap and sensor locations on the body are also explored.

Next, two factors that differentiate AR using wearable sensors from using a mobile phone are studied: the possibility of a mobile phone being carried (1) in an unknown and previously untrained location on the subject's body, and (2) with any orientation.

Key findings presented within the thesis include: (1) Acceleration derived features perform better than features of the other two sources. (2) Using acceleration features alone performs only marginally worse than using features derived from all three sources. (3) Orientation derived features have the second highest success-rate but require the lowest sampling frequency of features derived from the three sources. (4) Accuracy is affected by location and number of sensors. The wrists yield the highest overall performance for the activities studied. A depreciating returns relationship exists between accuracy and number of sensors used. (5) Of the body locations and activities studied, it is possible to identify the location on which the sensor is carried without knowing the subject's activity. (6) Reorienting the data from local to global coordinates ameliorates the decrease in success-rate when the sensor's orientation is transient. However, it incurs a marginal decrease when the orientation is fixed. (7) Reorienting the data using orientations obtained from an IMU results in higher success-rates than using accelerations. (8) Highly confusable activities are found to have similar gross motor movements.

Signature	Witness	Date
		4/8/2015

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed ……………………………………………………………

Date ……………………………………………………………

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed ……………………………………………………………

Date ……………………………………………………………

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ……………………………………………….............

Date ……………………………………………….............

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgment is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the projects design and conception or in style, presentation and linguistic expression is acknowledged.

Signed: ....

Umran Azziz Abdulla

Date: 27-02-2015

Acknowledgements

I would like to thank the following people (in no particular order):

1. Khushi – my better half in every sense of the phrase – who supported me through thick and thin as I obsessively pursued this research.
2. Dr Ken Taylor – whose unconventional way of looking at things taught me to think and question in ways that I did not think was possible before meeting him.
3. Dr Michael (Spike) Barlow – whose wisdom in dealing with life I have come to greatly admire and even emulate.
4. Dr Teddy Mantoro – for encouraging me to pursue a PhD.
5. My mom – whose strength and patience never ceases to amaze me.
6. My in-laws – for accepting me into a family that only ever existed in my dreams.
7. My dad and the rest of my family – for all the support they have given me over the years.
8. The members of the Virtual Environments & Simulation Lab (VESL) – for the support and friendship during my candidature.
9. The subjects – who gave up their time to participate in this research.
10. The University of New South Wales (UNSW), the Australian Defence Force Academy (ADFA) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO)– for providing the resources necessary to conduct this

research.

11. All those who directly or indirectly helped me undertake this research.

12. All those who directly or indirectly helped me get to this point of my life.

It’s so much easier to suggest solutions when you don’t know too much about the problem.

Malcolm S. Forbes,

The Sayings of Chairman Malcolm: The CapitalistâĂŹs Handbook (1978)

One accurate measurement is worth a thousand expert opinions.

Grace Hopper (source undetermined).

Abstract

Considerable research has been performed exploring Activity Recognition (AR) using wearable sensor nodes such as smart phones that incorporate accelerometers, gyroscopes and magnetometers.

This thesis presents an empirical evaluation of AR using on-body sensors. It studies the recognition of 22 Activities of Daily Living using either wearable sensors that have fixed locations on the subject's body or smart phones carried by the subject. The differences between the two sensor configurations are explored as well as parameters for computationally efficient and accurate AR.

Initially, fundamental classifier settings that impact AR accuracy are explored. These include evaluating the performance of acceleration, rotational velocity and orientation derived features (the three are referred to as *sources*). In addition, minimum sampling frequencies, window size, window overlap and sensor locations on the body are also explored.

Next, two factors that differentiate AR using wearable sensors from using a mobile phone are studied: the possibility of a mobile phone being carried (1) in an unknown and previously untrained location on the subject's body, and (2) with any orientation.

Key findings presented within the thesis include: (1) Acceleration derived features perform better than features of the other two *sources*. (2) Using acceleration features alone performs only marginally worse than using features derived from all

three *sources*. (3) Orientation derived features have the second highest success-rate but require the lowest sampling frequency of features derived from the three *sources*. (4) Accuracy is affected by location and number of sensors. The wrists yield the highest overall performance for the activities studied. A depreciating returns relationship exists between accuracy and number of sensors used. (5) Of the body locations and activities studied, it is possible to identify the location on which the sensor is carried without knowing the subject's activity. (6) Reorienting the data from local to global coordinates ameliorates the decrease in success-rate when the sensor's orientation is transient. However, it incurs a marginal decrease when the orientation is fixed. (7) Reorienting the data using orientations obtained from an IMU results in higher success-rates than using accelerations. (8) Highly confusable activities are found to have similar gross motor movements.

List of Publications

1. Abdulla, U. A., Taylor, K., Barlow, M., and Naqshbandi, K. Z. (2013). Measuring Walking and Running Cadence Using Magnetometers. In *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pages 1458âĂŞ1462. IEEE.
2. Taylor, K., Abdulla, U. A., Helmer, R. J., Lee, J., and Blanchonette, I. (2011). Classification With Smart Phones For Sports Activities. In *5th Asia-Pacific Congress on Sports Technology (APCST)*, pages 428âĂŞ433.

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
Accel.	Accelerometer or Accelerations
ADL	Activity of Daily Living
BN	Bayesian Networks
BSN	Body Sensor Network
CDF	Cumulative Density Function
FFT	Fast Fourier Transform
Gyro.	Gyroscope
HMM	Hidden Markov Model
IQR	Inter-quartile Range
KSOM	Kohonen Self Organizing Map
Orient.	Orientation or Orientations
PDF	Probability Density Function
RFID	Radio Frequency Identification
RMSE	Root Mean Squared Error
RSSI	Received Signal Strength Indication
Rot. Vel.	Rotational Velocity or Rotational Velocities
SSE	Sum Squared Error
S.D.	Standard Deviation

Glossary

The following are definitions of terms used within this thesis. Please note that the definitions of some of the terms used within this thesis might differ from the common definition of the terms.

Activity classification : The process of identifying an activity from a given closed set of activities by using a feature-vector that summarises a window of data captured during that activity.

Activity classification accuracy : The accuracy of the results of the activity classification process, measured as a percentage of the correctly classified feature-vectors out of the total number of feature-vectors.

Activity-location : The conjunction of data from a particular activity and data from a particular body-location. For example, all acceleration data captured from body-location L of all subjects doing an activity A is said to be acceleration data of activity-location (A, L).

Activity recognition : The overall process of identifying the subject's activity and the delivery of the activity information to the user, including signal processing performed before classification, the activity classification task itself, and any processes that might be applied to the activity information so to present consistent and coherent information captured about and of the user's activities. Within this thesis, a distinction is drawn between activity recognition and activity classification. Activity recognition refers to the overall process by which a system identifies and presents activity information while activity classification refers to the specific task of using a classifier to identify the activity that was occurring when the data that resulted in a feature-vector was gathered.

Activity recognition accuracy : The accuracy of the final activity information presented to the user. No section of this thesis attempts to measure activity recognition accuracy.

Body-location : A location on a subject's body.

Carry location : The location on the subject, that the subject carries his or her smart-phone.

Feature-extraction : The process of applying a feature-set to a window of sensor data to produce feature-vectors.

Feature-set : A set of functions that take a single window of accelerations, rotational velocities or orientations, and produce a single feature-vector. Within this thesis, a feature-set only consists of the functions described in section 3.5

for either Bao and Intille's feature-set or Kwapisz et al.'s feature-set. It does not include any processing that is/can be performed prior to the applications of the functions described in section 3.5. Such processing performed prior to the application of the functions to the signal, such as signal filtering and downsampling, is refered to as preprocessing. This distinction is important within this thesis because it allows for the analysis of the impact of each task (preprocessing or feature-selection) individually.

Feature-vector : A sequence of numbers, of a fixed length and order, that summarise characteristics of a window of accelerations, rotational velocities or orientations (i.e. a *source*), and that are used to train or test a classifier. Feature-vectors of accelerations, rotational velocities or orientations can be combined to form larger feature-vectors.

Monitor : A single module, containing an accelerometer, gyroscope and magnetometer that can be mounted on any location of the subject's body to record the accelerations, rotational velocities and orientations of the subject at that given location.

Preprocessing : Any processing that is carried out on sensor data before feature-extraction.

Source [1]: Either one of accelerations, rotational velocities or orientations. The three form the base source of sensory information on which activity recognition is studied in this thesis.

Window : The data gathered within a specified time period.

[1]The word is italicised because orientations are not so much a new source of sensory information as a result of combining accelerometer, gyroscope and magnetometer values. This is done in the IMUs used. While the algorithms used in IMUs are well known, the parameters of the algorithm can vary from one IMU implementation to another. However, orientations are real physical properties of objects and can be measured using other techniques other than IMUs (e.g. using motion capture cameras)

1 Introduction

1.1 Overview

Activity recognition (determining what activity a user is undertaking at the current moment) is an interesting research area for a variety of reasons including:

1. Ambient-assisted living where activity recognition is used to assist independent living and aging in place.
2. Pervasive mobile computing where activity recognition is used to infer the activities of the user using a mobile device and either have the device respond to the context (e.g. by adjusting the ring volume or changing the phone to vibration) or providing suggestions of similar activities and locations the user might enjoy.
3. Context-aware computing where systems installed in particular environments (such in a hospital or factory) aim to identify contexts and proactively provide services relevant to the context or assist in the current activity.
4. Surveillance-security where activity recognition is used to address threats to safety and security.

Advances in sensor technology and computation have allowed for better and smarter systems that can deduce user(s) activities. Activity recognition can be classified based on where the sensors are mounted (on-body sensors or dense sensing in intelligent environments), approach to understanding the problem (data-driven or knowledge-driven) and algorithms used (e.g. computer vision-based, statistical modelling etc.) (Chen et al., 2012).

The research work in this thesis can be termed as:

On-body sensor-based: The sensors the research work has made use of are worn on the subject's body. In particular, the sensors are inertial measurement units (IMUs) that measure the subject's inertial movement.

Data-driven and knowledge-driven approach: The approach used to understand the problem is both data-driven and knowledge-driven. Data of the activities of interest was gathered and used to model and recognise when the activities are occurring. In addition, frequent references and validations are made using extant knowledge about the kinematics of body movements and observations made during data gathering.

Statistical modelling: This research applies statistical methods to study body movements and develop statistical models that represent those movements.

This research work analyses the impact of several parameters on the classification of Activities of Daily Living. The impact of the parameters on activity classification success-rates is studied and conclusions are drawn about the relationships between the parameters and success-rates and the roles the parameters play or should play in activity recognition. The impact of the following parameters on activity classification success-rates are analysed:

1. The sampling frequency.
2. Using either accelerations, rotational velocities, orientations or a combination of all three as the source on which to extract feature-vectors.
3. The window length and window overlap to use while using the sliding window approach to segment data.
4. The locations on the user to mount sensors on and the number of sensors to mount on the user.

5. Models of training and testing models (as given by Lockhart and Weiss (2014)) and the role of inter-subject variation within the models.
6. Sensor orientation transiency relative to the body-location mounted on and the impact on activity recognition accuracy.
7. Application of data reorientation to convert data from local sensor coordinates to world coordinates.
8. Inter-location variation and the consequential ability to distinguish where the sensor is mounted on due to the variation.

In addition, several analyses are performed so as to obtain higher activity classification success-rates and more efficient activity recognition systems. The analyses are:

1. An analysis on the impact of sampling frequency on activity classification success-rates using a particular feature-set to sample more efficiently.
2. An analysis of how much mutual confusion exists between different activities, which activities are most highly mutually confused using a particular classification algorithm, and a way to visualise the mutual confusion existing within the classification results.

1.2 Motivation

While considerable work has been carried out towards application of on-body inertial sensors in activity recognition, much of the work has been constrained to accelerometers. Recent work has started to include gyroscopes and magnetometers. The narrow focus that has existed till recently is possibly because of the unavailability of these sensors to the activity recognition research community, but this is changing as sensors become less expensive and (mobile) computation power is more readily available and less power consumptive.

Another recent area of activity recognition research is the use of smart phones to recognise and record users' physical activities. This involves making use of the sensors available on smart phones (which currently include accelerometers, magnetometers, gyroscopes, light sensors and GPS) to identify the user's activities.

High accuracies have already been attained in activity recognition using on-body inertial sensors. However, this has mostly been focused toward identifying simple activities. While many simple activities exist in people's day-to-day lives, there are many more complex activities than there are simple activities. This work studies several parameters that influence the classification success-rate of activities that people perform in their day-to-day lives in the hopes that if the behaviours of these parameters are better understood, it would result in more accurate and more efficient activity recognition systems.

1.3 Scope of research

This research limits itself to the set of just over twenty activities gathered that include food preparation, household cleaning and office work. In addition, the research is limited to the use of statistical features extracted using a sliding window technique. In the thesis, two feature-sets are selected and each analysis is repeated for the two feature-sets in order to strengthen the conclusions made from the results. In terms of classifiers, the analysis is performed using decision trees because, via a pilot study, they were found to be the best classifier for the feature-sets. No analysis is presented in the thesis itself on the performance of different classifiers.

Due to the data gathering process, two distinct sets of data are available: one for walking and running that was gathered with a set of three monitors only; and one dataset containing a further 20 (more sophisticated) Activities of Daily Living that was gathered using a set of six monitors. Where possible the analysis is repeated for both the three monitor set and all 22 activities, and six monitor set with 20 activities.

1.4 Original contributions

The original contributions made in this thesis are:

1. Feature-vectors extracted from accelerations result in higher activity classification success-rates than feature-vectors extracted from either rotational velocities or orientations.
2. Feature-vectors extracted from accelerations alone result in marginally worse activity classification success-rates than using feature-vectors extracted from the three *sources* combined.
3. Feature-vectors extracted from orientations have the second highest success-rate of feature-vectors derived from any of the three sources, but require the lowest sampling frequency.
4. Accuracy is affected by the body-location on which sensors are mounted and the number of different body-locations on which the sensors are mounted. The wrists yield the highest overall performance for the activities studied. A depreciating returns relationship is observed between the accuracy obtained and number of sensors used.
5. Of the body-locations and activities studied, it is possible to identify the location on which the sensor is carried without knowing the subject's activity.
6. Reorienting the data from local to global coordinates ameliorates the decrease in success-rate when the sensor's orientation is transient. However, it incurs a marginal decrease when the orientation is fixed.
7. Reorienting the data using orientations obtained from an IMU results in higher success-rates than Reorienting the data using the current best acceleration-based reorientation method in literature.
8. Highly confusable activities are found to have similar gross motor movements.

1.5 Thesis organisation

The thesis is organised as follows: first a review of literature is given; an overview of the methodologies used in the analysis; after which four analysis chapters are then given. Each analysis chapter is divided into sections that represent one experiment. In each experiment, the methodology used in analysing the data is given followed by the results obtained and a discussion of the possible reasons behind any observations made in the results and implications of the results.

The four analysis chapters are as follows. The first chapter analyses the impact of data capturing and feature-extraction approaches upon activity classification. The second chapter analyses the differences in the activity classification success-rates obtained using the two feature-sets studied, the impact of the number of monitors used per subject and location of monitors on the subject's body on the activity classification accuracy. The third chapter studies inter-subject and inter-activity variability and their impact on activity classification accuracy. In addition, the mutual confusion rates between activities is also studied in order to identify which activities are highly confused with each other. The forth chapter studies smart-phone based activity recognition and in particular analyses the impact of random monitor rotations (relative to the body-location on which phone is carried) and the identification of the body-location on which the phone is carried.

Each analysis chapter is concluded with a summary of the chapter. In addition, an additional summary of the whole thesis is found in the conclusion chapter.

2 Background and literature review

This chapter will present a break-down of literature on activity recognition using on-body inertial sensors. In addition, literature is presented on recent interesting and relevant time-series data-mining techniques, human kinematics, and an overview of the current state of commercial on-body activity trackers.

2.1 Definitions and overview

Context is any information that can be used to characterise the situation of an entity, where an entity can be a person, place or physical object (Dey, 1999). The ability of a system to sense various states of it's environment and itself is termed as context-awareness (Pascoe, 1998). Pascoe (1998) presents four context-awareness capabilities a system can support:

1. Detecting context and presenting it to the user so as to enhance the user's sensory system.
2. Adapting the system's behaviour to suit the current context.
3. Allocating and using resources that are accessable from or relevant to a given context.
4. Augmenting the environment with contextual information.

Dey (1999) opines that certain types of context are in practice more important than other types. These are: location, identity, activity and time. Activity-awareness is a subcategory of the context-awareness research area. The goal of it is to identify the userâĂŹs actions and activities from sensor observations.

Kim, Helal, and Cook (2010) explains the the goal of activity recognition is the accurate detection of activities based on predefined activity models. To perform this, models of the activities need to be built. Activity pattern discovery is the process by which low-level sensor data is analysed to find previously unknown activity patterns. Bannach, Amft, and Lukowicz (2008) explain that activity recognition systems are developed in 2 main phases. The first phase involves the design of the recognition method such as sensor setup, selecting features and classifier. The second phase involves implementing the selected algorithms in the appropriate devices.

Gu, Wu, Wang, Tao, and Lu (2009) categorise instances of Activities of Daily Living (ADLs) with relation to other subject's activities into:

1. Activities that a person does individually and with no connection to others,
2. Activities two or more people do individually but simultaneously (e.g. having dinner together, watching TV together, etc.),
3. activities in which two or more people cooperate to achieve a shared goal (e.g. preparing a meal together), and
4. Activities in which two or more people perform individually but whose goals conflict (e.g. one person wants to shower while another wants to use the toilet, one person wants to play computer games while another wants to check mail).

In addition, activities can also be grouped into simple or atomic and complex activities. Dernbach, Das, Krishnan, Thomas, and Cook (2012) define atomic activities as those that consist of a single repeated action whereas complex activities are composed of a series of multiple actions.

Activity recognition systems either use sensors placed on the user's body (on-body sensors) which are mostly worn as part of the subject's clothing (wearable sensors) or make use of sensors installed in the subject's environment (intelligent environment) possibly making use of sensors attached to items that the subject manipulates (dense-sensing). Dernbach et al. (2012) explains that one disadvantage of using sensors installed in the environment is the initial cost of installation.

On the other hand, wearable sensors require a daily effort by the subject to put the sensors on. One possible option is to use the user's smart phone instead of wearable sensors. Using the smart phone as the sensing node avoids these issues but introduces other challenges and limitations such as limited processing power, limited battery life, variable and expensive communication bandwidth and a requirement to still operate as a phone (Taylor, Abdulla, Helmer, Lee, & Blanchonette, 2011). Dernbach et al. (2012) explains that current smart phones are equiped with many sensors including GPS sensors, microphones, cameras, light sensors, proximity sensors, inertial sensors (accelerometers and gyroscopes) and direction sensors (magnetometers). However, while smart phones are equipped with many sensors, the majority of users only carry one smart phone at a time. Hence, an activity recognition system that only depends on sensors on a smart phone would at most only monitor motions of one location on the subject's body.

In order to infer the userâĂŹs current activity, activity recognition systems involve three main components: a sensing component that reads raw environment data from sensors; a feature extraction component that processes the raw data to extract a set of features from the data; and a classification component which uses the extracted features to determine the userâĂŹs activity (Choudhury et al., 2008).

The next sections cover research concerning different applications of on-body activity recognition, research performed in each of the three mentioned components of activity recognition systems, an overview of the current state of commercial activity trackers and finally other related research.

2.2 Applications of on-body activity recognition

Activity-recognition has initially been applied to detecting specific activities in narrow contexts and only more recently to detecting more general daily activities. Specific applications of activity-recognition include:

Recommending leisure activities

Bellotti et al. (2008) present a system called *Magitti* that infers the user's current activity based on his or her location, current time, weather, user input, viewed apps, user's calendar and emails sent or received. Their system recognises the user's current and future leisure activities and recommends suitable leisure locations.

Personal emergency systems

Mathie, Coster, Lovell, and Celler (2004) survey work performed using accelerometers to monitor gait, sit-to-stand transfers, postural sway, falls, energy levels and identifying subject's movements. Mathie, Coster, et al. (2004) then propose a system that integrates these components for monitoring subjects in order to access functional status in an unsupervised free-living environment.

Healthcare

Tentori and Favela (2008) performed a nine month survey in a hospital where they shadowed a number of hospital staff noting their actions as they were executed. They proposed a system that assists nurses in a hospital to care for patients using sensors connected to hospital devices (e.g. on urine bags), a bracelet that acts as an indicator of particular user-selected states, and an application on the nurse's smart phone that provides further details on the patient's state and also allows some control of the system (e.g. selecting what state(s) the bracelet should alert the nurse of).

Assembly-line workers

Stiefmeier, Roggen, Ogris, Lukowicz, and Tr (2008) describe a system developed to monitor assembly-line workers in car manufacturing as they perform quality assurance checks. Sensors (which include several IMUs and force resistive resistor straps) are incorporated onto a jacket worn by workers to serve as sensing nodes. Other sensors are fixed onto tools and the work area. Together these sensors recognise the activity the worker is performing and detect when a step in the checking procedure is missing and allow the worker to enter detected faults directly into the database instead of noting it down on paper.

Within Specific sports

The following are examples of the application of activity recognition to specific sports by using on-body inertial sensors found in activity recognition literature.

1. *Professional skiing*

   Michahelles and Schiele (2005) use force sensors, triaxial accelerometers and gyroscopes, radar units and an infra-red distance sensor to improve the relationship between trainers and athletes by allowing them to share their observations and impressions.

2. *Golf*

   Ghasemzadeh, Barnes, Guenterberg, and Jafari (2008) describe a golf training system which incorporates wearable motion sensors to obtain inertial information. The system analyses four major segments of a golf swing: takeaway, backswing, downswing and follow-through. It then provides feedback to the athlete on the quality of the movements.

3. *Rowing*

   King et al. (2009) present a prototype system that monitors the kinematic of the femur and the lower back during rowing. The data is collected from inertial sensors attached to the rower. The data is then used to determine the rotation of the lower back and femur, which is in turn used to identify some common poor rowing techniques.

2.3 The sensing component

This section addresses research conducted in areas related to the sensing component of an activity recognition system (Choudhury et al., 2008).

Sensors analysed

Accelerometers are by far the most common on-body sensors used in activity recognition (Chen et al., 2012), although other sensors have been used in some studies. Such sensors include gyroscopes (Ghasemzadeh et al., 2008), signal strength (RSSI) (Quwaider & Biswas, 2008), and GPS, ECG, respiratory effort sensor and oximeter (Ermes & Juha, 2008). Wang et al. (2009) present a system which uses the current user's state and selects the appropriate sensor from the sensors available on smart phones to update the user's current location, activity and environment while minimising the impact on the battery life of the phone. The sensors employed include accelerometers, light sensors, microphones, GPS, Wi-Fi scans and Bluetooth scans.

Performance of a combination of inertial sensors

Kunze and Lukowicz (2008) present a system that makes use of both accelerometer and gyroscope to provide accurate activity classification that is tolerant to sensor displacement within a single body part (e.g lower arm). They decompose acceleration readings at a particular body-location into three: gravity component with respect to orientation of the sensor, sensor translation and sensor rotation. Only the last is sensitive to sensor displacement on the same body part. They propose that if a system had both accelerometer and gyroscope sensors, then to make use of the gyroscope when the rotational component is dominant, otherwise to make use of the accelerometer.

Amft and Tröster (2008) use a number of IMUs mounted on the upper and lower arms to recognise arm movements while eating. In addition to IMUs they also use microphones and electromyograms to detect chewing and swallowing. The goal of the research was automatic diet monitoring. Good success rates were achieved for recognising food intake arm movements and two food groups based on chewing sounds.

Yuwono, Su, Moulton, and Nguyen (2012) use a chest-mounted IMU to detect walking and measure cadence. Their method involves obtaining the orientation of the sensor by using a complimentary filter that combines gyroscope data and accelerometer data. The pitch angle is then extracted from the sensor's orientation. The pitch data is used in further signal processing to extract cadence and determine whether the activity is walking or not.

Dernbach et al. (2012) analysed the use of accelerations and orientations gathered from smart phones carried by subjects in recognising simple and complex activities performed by the subjects. Their results found that between using accelerations alone and accelerations combined with orientations, combining both accelerations and orientations lead to better results than when accelerations alone. Using orientation data in addition to accelerations resulted in 10-12% increase in activity classification success-rates than was obtained by using accelerations only.

The closest analysis performed with relation to the impact of different types of inertial sensors on activity classification accuracy is Shoaib, Bosch, Incel, Scholten, and Havinga (2014). Shoaib et al. analyse the application of various motion sensors available on smart phones for recognising stationary (sitting and standing), ambulatory activities (walking on a flat surface, walking up stairs, walking down stairs, jogging, running) and biking. They conclude that, with the exception of the magnetometer, the other sensors (accelerometers and gyroscopes) have the potential to deliver positive classification performance on an individual sensor basis, depending on the activity and body-location. They also conclude that the sensors need only be combined in the case when both sensors are not performing well. It is important to note that even though Shoaib et al. express the impact of orientation changes on the data, other than adding an additional set of features based on the magnitude of the accelerometer vectors, they do not perform any other preprocessing to mitigate the impact of orientation changes or even deal with the possibility that the system users might carry their phones in different orientations from the orientations featured in the training data. Among the observations they made include:

1. The gyroscope performs better than the accelerometer for walking upstairs and downstairs, when the phone sensor is carried in the thigh pocket and belt positions. This could possibly be due to rotations occurring while walking up stairs and down stairs that affect the accelerometer data but are measured by the gyroscope data.
2. The accelerometer performs better than the gyroscope for identifying sitting and standing (stationary activities or postures). This is because identifying these postures is really about identifying the orientation of the sensor, and since gyroscopes measure angular velocity, a gyroscope would read the same values despite the orientation the sensor is in.
3. Both the gyroscope and the accelerometer recognised biking, jogging and walking well, although the accelerometer performs better.
4. The linear acceleration sensor performs as well as the accelerometer, however it is better than the accelerometer for the upper arm position and poorer at the belt and thigh pocket locations. This could possibly be an effect of the gravity

component in defining the orientation of the sensor in various activities. It is worth mentioning that the linear acceleration sensor is a pseudo-sensor on Android smart phones that provides gravity subtracted accelerometer values.

It is important to note that while Shoaib et al. analysed accelerations and rotational velocities, and Dernbach et al. studied accelerations and accelerations combined with orientations. Shoaib et al.'s analysis did not include orientations and Dernbach et al.'s analysis did not include rotational velocities or independent orientations (without accelerations). No studies were found that studied activity classification rates achievable by orientations only, or compared activity classification success-rates achievable by accelerations, rotational velocities and orientations.

In addition, it is worth noting that Dernbach et al.'s research is focussed on activity recognition using a smart phone. In Dernbach et al.'s data collection, each subject was allowed to carry the phone in the location and orientation that the subject prefered. As a result, it is likely that the phone's orientation was different from one subject to the next, or even at different times for the same subject. While this is important for activity recognition systems that use smart phones as sensing nodes, it is highly likely that the changes in phone orientations could be negatively impacting activity classification success-rates. This would differentiate the results obtained using a smart phone as a sensing node, to those obtained using inertial sensors which are worn in a fixed orientation relative to the body-location the sensor is mounted on.

In addition, Shoaib et al.'s research is also focussed on activity recognition using a smart phone. While Shoaib's data gathering, the locations and orientations of the phones used as sensing nodes are fixed relative to the body-location, the data obtained from each body-location is processed and classified independent of the other body-locations. However, it is unclear how accelerations, rotational velocities and orientations would perform in an activity recognition systems that combines information from multiple body-locations to recognise the user's current activity.

Hence, it would be valuable to understand which of three (accelerations, rotational velocities and orientations) results in higher activity recognition success-rates and compare the individual success-rates of the each of the three *sources* to those of the three *sources* combined in an activity recognition system that utilises multiple inertial sensors mounted on different body-locations and whose orientation and location is fixed relative to the body-locations the sensors are mounted on. This setup is likely to result in higher activity classification success-rates than when either: only data from a single body-location is used; or when the orientation of the sensor relative to the body-location it is mounted on is transient.

Sampling frequency

Several sampling frequencies have been used in the literature. These include 10 Hz (Khan, Lee, & Kim, 2008; Sun, Zhang, Li, Guo, & Li, 2010), 15 Hz (Baek, Lee, Park, & Yun, 2004), 20 Hz (Bieber, Voskamp, & Urban, 2009; Ermes & Juha, 2008; Kwapisz et al., 2011; Quwaider & Biswas, 2008), 22 Hz (Ghasemzadeh et al., 2008), 32 Hz (Lee, Park, Hong, Lee, & Kim, 2003), 36 Hz (Yang, 2009), 45 Hz (Mathie, Celler, Lovell, & Coster, 2004), 50 Hz (Henpraserttae, Thiemjarus, & Marukatat, 2011; Keally, Xing, & Pyles, 2011), 200 Hz (Hanai, Nishimura, & Kuroda, 2009). However, little work has been performed in analysing sampling frequency requirements for activity recognition.

Bieber et al. (2009) explain that human muscle movements are controlled by information transferred by nerves and the response time of humans depends on the kind of signal (acoustic vs optical), muscle temperature, psychological and physiological constitution, as well as external influences such as drugs, alcohol and nicotine. Further, they explain that the average approximate response time to optical stimuli is 220 ms (approximately 4.5 Hz), while acoustic stimuli ranges from 10 to 13 Hz. A reflex action however occurs without brain processing and occurs at approximately 16 Hz. Therefore a sampling frequency of 32 Hz is sufficient to sample body movements.

Maurer, Smailagic, Siewiorek, and Deisher (2006) experienced increase in success rates as they increased sampling frequency to about 15 - 20 Hz where the success rates stabilised and after which only marginal improvements were obtained. Maurer et al.'s analysis looked at sampling frequencies in the range [0Hz,50Hz] for six stationary or ambulatory activities (sitting, standing, walking on a flat surface, walking up stairs, walking down stairs and running) using data captured from six subjects using six 2D accelerometers mounted on the left wrist, belt, necklace, right trouser pocket, shirt pocket and bag.

Maurer et al.'s study only analysed activity classification success-rates obtained from feature-vectors extracted from accelerations. However, it in unclear whether Maurer et al.'s findings about the relationship between sampling frequencies and activity classification success-rates also apply to feature-vectors extracted from rotational velocities and orientations. In addition, Maurer et al.'s study only included stationary and ambulatory activities. It is unclear whether the findings would apply to other types of Activities of Daily Living, for example, household cleaning activities, laundry activities and desktop activities.

Mobile phone carrying locations

Bieber et al. (2009) explain that a fixed location on which the users would have to place their smart phones for the application to work would not fit users' everyday life behaviour. They refer to a survey taken by Nokia (Cui, Chipchase, Ichikawa, & Tokyo, 2007) of 1,549 participants from 11 cities in 4 continents. The survey characterises the location in which users usually carried their mobile phones. Table 2.1 gives locations where people carry their phones and the percentage of people that carry their phones in those locations for men and women according to Cui et

Table 2.1: Percentage of men, women and combination of men and women that carry their mobile phones at particular locations on the body.

Location	Men (%)	Women (%)	Men & Women Combined (%)
Trousers/Skirt	60.10	16.42	38.26
Bags	10.10	61.06	35.58
Belt case/clip	13.79	0.81	7.30
Hands	3.45	9.09	6.27
Upper-body	8.25	2.17	5.21
Other	1.97	6.11	4.04
Not with me	2.09	1.90	2.00
Neck	0.25	2.44	1.35

al. (2007). An additional column has been added showing combined percentages (assuming equal numbers of men and women).

With reference to table 2.1 the highest combined percentage of users place their phones in the trouser or skirt location followed by bags. Another interesting fact is that a majority of men (60.10%) place their phones in their trouser pockets while a majority of women (61.06%) place their phones in bags.

Best body-location to monitor Activities of Daily Living

The closest research found that analysed the impact of the location on which a sensor was mounted on, upon the activity classification success-rate, is that of Henpraserttae et al. (2011); Keally et al. (2011); Maurer et al. (2006) and Shoaib et al. (2014).

Maurer et al. noted that features derived from a 2D accelerometer mounted on the subjects' bag resulted in higher success-rates than those derived from accelerometers mounted on the left wrist, belt, necklace, right trouser pocket and shirt trouser pocket. The research included walking, running, sitting, standing, and ascending and descending stairs for six subjects. It is interesting to discover which body-location, results in the best success-rates in a few more thorough contest–greater range of activities, larger pool of participants, larger range of sensor sources.

Keally et al. analysed the success-rates achieved by accelerometers mounted on different locations on the body in recognising postures (sitting and standing), ambulatory activities (walking on a flat surface, walking up stairs and walking down stairs), writing on a whiteboard, typing on a keyboard and shaking hands. They mounted accelerometers on both sides of the body on the following locations: above the ankle, above the knee, the hip, the wrist, above the elbow and on the shoulder. They noted that the activity classification success-rates obtained from data captured from accelerometers mounted on the legs was significantly lower than success-rates obtained from data captured from accelerometers mounted on the upper body. This is because some of the activities (writing on a whiteboard, typing on a keyboard and shaking hands) were not recognised well using data captured from sensors mounted on the subject's leg. Between the left and right side of the body, they noted that shaking hands was more accurately identified using data from sensors mounted on the right arm, and typing on a keyboard and writing on a whiteboard was more accurately identified using data from sensors mounted on the left arm due to the more discriminative posture of the left arm in these activities. For sensors mounted on the leg, they noted that the accuracy at which data captured from the sensors could identify sitting, standing, walking on a flat surface, walking up stairs and walking down stairs, was similar enough that they recommended the use of only one sensor instead. However, it should be noted that Keally et al.'s data was gathered from one subject only.

Although the research presented by Henpraserttae et al. is primarily about analysing the impact of accelerometer signal transformations from local to global coordinates on smart-phone-based activity classification, part of their work analysed the performance obtained from data captured at different body-locations. The locations tested only included locations on which a phone could be carried: shirt-pocket, trouser-pocket and waist. They found that the waist results in the highest success-rate, followed by the shirt-pocket and trouser-pocket.

Although the research presented by Shoaib et al. is primarily about comparing success-rates obtained by sensor-types available on smart-phones (accelerometers, gyroscopes and magnetometers), their work also analysed activity classification success-rates of several locations on the body where phones are carried. The locations analysed are: left trouser pocket, right trouser pocket, belt position towards the right leg, upper right arm, lower wrist. Shoaib et al. conclude that the location that results in the best activity classification success-rates varies from activity to activity.

Number of body-locations monitored

Zappi et al. (2007) compared the number of monitors used to recognise activities of automotive assembly line workers and found that a monitor mounted on one body-location could at best yield a success rate of 50% rising to 80% for 3 body-locations and 98% for 57 body-locations. This indicates that success rate increases with the number of body-locations being monitored but the relationship is not linear.

It is interesting to note that this research covers a specialised situation: that of an automotive assembly line. There has, to date, been no communication in the research literature, concerning the relationship between the number of inertial sensors mounted on a subject and the success-rates of recognising Activities of Daily Living. It is interesting to find out whether this same relationship would be observed for Activities of Daily Living. Since the characteristics of motions performed in Activities of Daily Living are likely to be different from the motions performed on an automotive assembly line, it is likely that this analysis would result in insights on the number and locations of the body-locations that need to be monitored in order to achieve high activity classification success-rates for Activities of Daily Living.

2.4 Feature-extraction component

This section covers research performed in areas related to the feature-extraction component of an activity recognition system (Choudhury et al., 2008).

Windowing techniques

Before features are extracted, the time-series in on-body activity recognition are commonly divided into windows. Features are then extracted from each window and used for classification (Preece, Goulermas, Kenney, & Howard, 2009).

Preece, Goulermas, Kenney, Howard, Meijer, and Crompton (2009) explains that the three windowing techniques observed in activity recognition literature are: sliding windows, event-based windows and activity-based windows. Event-based windowing uses specific events in the time-series (such as the detection of a heel strike in accelerometer data) to segment the data. Activity-based windowing depends on identifying transitions between activities before identifying the activity itself. Sliding windows use windows of equal sizes, and that are uniformly distributed in time. Of the three, sliding windows is the most commonly used windowing technique due to its simplicity (Preece, Goulermas, Kenney, Howard, Meijer, & Crompton, 2009).

Windowing Lengths

Several different sliding window lengths have been used in the literature ranging from 1 second (Henpraserttae et al., 2011; Sun et al., 2010), 2 seconds (Baek et al., 2004) to 10 seconds (Keally et al., 2011; Kwapisz et al., 2011; Yang, 2009).

In their analysis of the performance of different sliding window lengths, Huynh and Schiele (2005) analysed window lengths of 0.25, 0.5, 1, 2 and 4 seconds and found that no one window length performs best for all activities. The one second window performed best when the activity being tested was walking or jogging, the

2 and 4 second windows performed best for skipping and hopping, while the 0.25 and 0.5 second windows performed best for standing. No deeper reasoning was provided as to why these differences occurred. Huynh et al. used a single monitor mounted on the shoulder strap of the subjects' backpacks. The monitor included a 3D accelerometer, an audio sensor (no further details given), a temperature sensor, a light sensor, a humidity sensor, a barometric sensor and a magnetometer.

It is important to note that the analysis conducted by Huynh et al. included various ambulatory activities and riding the bus. It is likely that different window lengths are necessary for other activities. A valuable contribution could be made by gaining a better understanding of the relationship between window lengths and activity classification success-rates obtained for Activities of Daily Living like cleaning activities, cooking activities and desktop work (using a PC, writing, etc.).

In their analysis of the impact of window lengths on activity classification success-rates, Dernbach et al. (2012) found that shorter window lengths performed significantly better than longer window lengths. However, Dernbach et al. used a single sensor per subject which was carried in the body-location and orientation that the subject prefered (as is the case in smart-phone-based activity recognition). It is possible that the variation in sensor orientation and body-location have an impact in the classification results obtained. In which case, relationship might not hold in a case where multiple sensors are used per subject, and the location and orientation of the sensors relative to the body-location are fixed (as is the case with wearable-sensor-based activity recognition).

Contrary to Dernbach et al.'s finding, Patel, Mancinelli, Healey, Moy, and Bonato (2009) found that smaller window lengths resulted in lower activity classification success-rates. Patel et al. analysed the activity classification of gym activities (walking on a treadmill, riding a stationary bike and using an arm ergometer) and Activities of Daily Living (walking on a flat surface, walking up and down an inclined ramp, walking up and down stairs, folding laundry, sweeping the floor, using the bathroom, using the cafeteria and riding an elevator) from 15 subjects that were suffering from Chronic Obstructive Pulmonary Disease (OCPD). Their data was captured from 10 monitors (each monitor has a 3D accelerometer and 3D gyroscope) mounted on both arms (upper arm and lower arm), both legs (upper leg and lower leg), trouser pocket and on the sternum. They observed that increasing the window length beyond 6 seconds only resulted in marginal gains in activity classification accuracy.

Windowing Overlaps

The window overlap is the percentage or duration of a window that overlaps with adjacent windows. There has, to date, been no communication in the research literature, concerning the impact of window overlaps upon an activity classification algorithm's accuracy.

The most common window overlap found in activity recognition literature is 50% of the selected window length. A 50% window overlap was used by Bao and Intille (2004); Figo, Diniz, Ferreira, and Cardoso (2010); He, Liu, Jin, Zhen, and Huang (2008); Krishnan and Panchanathan (2008); Kunze, Lukowicz, Junker, and Tröster (2005); Preece, Goulermas, Kenney, and Howard (2009); Ravi, Dandekar, Mysore, and Littman (2005); Shoaib et al. (2014) and Sun et al. (2010). However, other window overlaps also exist in the literature review including: no overlap ((Kwapisz et al., 2011)), 20% ((Reiss, 2014)), 25% overlap ((Henpraserttae et al., 2011)), 33% overlap ((Lester, Choudhury, Kern, Borriello, & Hannaford, 2005)).

Larger window overlaps result in more windows hence more feature-vectors to train and test the classifier with. However, while this results in more data processing, it is unclear whether this results in higher success-rates or not. In an activity recognition system, the additional windows resulting from larger window overlaps result in additional feature-extract and classification. Hence, large window overlaps in activity recognition systems impact the computational resources of the system

Table 2.2: Frequency-domain features used in various publications

	Frequency-components	"start-to-end" amplitude	"peak-to-peak" amplitude	Power of frequency peak	Signal power at different frequency bands	Spectral entropy	Peak frequency	Energy	Power spectrum centroid	Frequency domain entrophy	Wavelet transform coefficients
Song and Wang (2005)											✓
Huynh and Schiele (2005)	✓					✓		✓			
Ermes and Juha (2008)				✓	✓	✓	✓				
Ghasemzadeh et al. (2008)		✓	✓								
Hanai et al. (2009)								✓		✓	
Yang (2009)									✓	✓	

and the power consumption of the system. It would be valuable to understand the relationship between the size of the window overlap and the activity classification success-rates obtained. This understanding could assist future researchers and activity recognition system developers to select the appropriate window overlap to use.

Features analysed

Among the activity recognition features analysed include time domain features and frequency domain features. Table 2.2 and table 2.3 show an overview of frequency-domain and time-domain features used in various activity recognition publications.

Carós et al. (2005) proposed a set of low complexity features for classifying ambulatory activities. Their proposed features rely on the computed energy of the

Table 2.3: Time-domain features used in various publications

	25% percentile	75% percentile	Autoregressive modelling coefficients	Correlation	Difference of angle of gravity from standing angle of gravity	Distribution of accelerations across multiple bins	Duration in current posture	Eccentricity	Inter-quatile range	Kurtosis	Mean	Mean acceleration magnitude	Mean of the absolute difference from the mean	Median	Probability rules based on tilt angle	Range	RMS power	Signal magnitude area	Signal magnitude vector	Skewness	Standard deviation/variance	Threshold of high-pass filtered signal	Tilt angle	Time between peaks	Zero-crossings
Lee et al. (2003)					✓											✓						✓			
Baek et al. (2004)								✓		✓	✓									✓	✓				
Carós, Chetelat, Celka, and Dasen (2005)																		✓	✓						
Huynh and Schiele (2005)											✓										✓				
Ermes and Juha (2008)	✓	✓								✓	✓			✓		✓				✓	✓				
Ghasemzadeh et al. (2008)											✓						✓				✓				
Khan et al. (2008)			✓															✓					✓		
Mathie, Coster, et al. (2004)							✓								✓								✓		
Hanai et al. (2009)				✓							✓										✓				
Wang et al. (2009)																					✓				
Yang (2009)		✓							✓		✓										✓				✓
Kwapisz et al. (2011)						✓					✓	✓	✓								✓			✓	
Sun et al. (2010)																									
Henpraserttae et al. (2011)											✓										✓				

signal to distinguish between static and dynamic activities and between walking and running, the repetitiveness of steps (refered to as harmonicity) to distinguish between ambulatory and other activities, and a feature computed from the double integration of the vertical acceleration to differentiate between walking on a flat surface from walking up or down stairs. The 3D accelerometer was assumed to be on a fixed location and orientation.

Dealing with possibilities of different sensor placements

In activity recognition systems that make use of a phone as a sensing node, the phone is likely to be placed on different body-locations at different times. Unlike wearable sensors that are worn on particular body-locations and fixed orientations relative to the body-location, there are many combinations of possible body-locations and orientations that a mobile device can be carried.

Kunze et al. (2005) decomposed sensor placement into three:

1. The body part on which the sensor is placed. This can be specific for some devices or can vary for mobile devices (e.g. smart phones).
2. The orientation in which the sensor is at any given time. Again this can be specific for some devices but can vary for others. Proposed methods that can be used to estimate the orientation of the sensor will be explained further in this section.
3. The exact position of the sensor on the body part. This is not solely an issue for mobile devices but any sensor which the user or subject puts on. The exact location on the body part is not guaranteed and Kunze et al. (2005) explain that simple calibration gestures are not sufficient to determine the exact location of the sensor on the body part. Kunze et al. (2005) propose a way of dealing with this issue by ignoring rotation dominated accelerometer readings and compensating with gyroscope data.

It should be noted that while 1 only happens when mobile devices are used as sensing nodes, 2 and 3 also happen for other wearable on-body sensors. As long as the subject puts the sensors on, the position might be inexact or inconsistent with previous wearings, and hence an algorithm that is not robust enough to deal with slight changes in orientation and positioning might suffer a decrease in performance. Kunze and Lukowicz (2008) explain that it is even possible for sensors embedded into tight clothing to be mispositioned by subjects not wearing the clothing as originally intended (e.g. subject might fold sleeves).

Particular to context-recognition systems that use a mobile sensing device, Kunze and Lukowicz (2008) explain that it is necessary to address the following location and orientation issues:

1. Body-location coverage: there should be enough body-locations covered by sensors to provide information for the activity recognition. The question that needs to be answered is, "Do the location(s) of the sensor(s) provide enough information to differentiate between activities?"
2. The sensor could be in any orientation: either the activity recognition algorithm should be orientation invariant or the system needs to be able to determine the sensor's orientation.
3. The sensor could be at any one of many body-locations: either the activity recognition algorithm should be body-location invariant, be able to deal with multiple locations or the system needs to be able to recognise where on the body the sensor is placed. Kunze and Lukowicz (2008) propose a way to deal with this issue by using a location invariant algorithm to determine when the subject is walking, then uses another algorithm to match the observed signal to the most likely body-location.

Based on the observation that there are only a limited number of locations that phones are carried and that in those locations there are some orientations that are more common, Sun et al. (2010) propose a meta-classifier that uses one classifier for each orientation-location. Even though they obtained good results with this method, running 32 classifiers (8 possible locations each with 4 possible orientation) as they did, would present a significant load on the constrained processing environment present on current smart phones. Furthermore, it is not clear what proportion of the phone-carrying population their algorithms would work with since their results are mostly tested on data gathered in a controlled lab environment. It is possible that there could be more orientations of a phone in a pocket than what is estimated by Sun et al. (2010) and it is not clear what results their algorithm would have with unaccounted for orientations.

Yang (2009) explains that even when these possible location-orientations are grouped, it is impractical to develop a separate inference model for each group of location-orientation. A solution was earlier proposed by Mizell (2003). He explains that the orientation of the device can be converted to a more general world orientation with one axis aligned to gravity (up-down or vertical axis), and two horizontal components, by making use of the fact that the accelerations measured by the device's accelerometer include both a static and a dynamic component. The static component is a result of the pull of gravity, while the dynamic component is a result of the subject's motions. Taking the average of the samples allocated over a period of time results in the gravity component. This can then be subtracted from the original samples to find the dynamic portion of the accelerations.

Extracting the gravity component this way assumes that no rotations occurred during the sampling window, which would work well for location-activity combinations without significant rotations (like chest locations while walking and running), but not for others (like hand locations while walking and running). Assuming a well calibrated accelerometer that is not free-falling we expect the magnitude of the gravity component to be $9.8m/s^2$. It can be shown that an effect of rotation is to lower this value (Kunze & Lukowicz, 2008). This would allow one to check whether the magnitude of the gravity vector is below the expected by a given threshold and throw away the segment if it were. Kunze and Lukowicz (2008) expand on this idea by proposing a method that determines which samples are dominated by rotation by making use of an accelerometer and a gyroscope. Samples dominated by rotation would have high rotational velocities and low acceleration. The opposite would hold for samples that are not dominated by rotation.

Rotating the accelerations based on the gravity vector gives a vertical and two horizontal components. Yang (2009) used the vertical vector and magnitude of the horizontal vector to extract features and perform classification. By taking this approach, Yang avoided dealing with each orientation independently and made the computation orientation-independent. Henpraserttae et al. (2011) extended this work by deriving the forward and sideways vectors from the horizontal vector by applying Principal Component Analysis (PCA).

However, this approach does not deal with the issue of multiple possible sensor locations. Henpraserttae et al. (2011) found that they achieved as little as 36% overall success rate (across all activities) when they trained their system with data obtained from the chest pocket and tested it against data obtained from the trouser pocket. Kunze and Lukowicz (2008) made a similar observation.

Kunze et al. (2005) propose an alternative approach to dealing with multiple possible sensor locations: deducing the location of the sensor. Since the patterns observed by the sensor are specific to both location and activity, Kunze et al. (2005) first decouple location from activity by using a location-invariant algorithm to ascertain that the activity is walking, then find the best matching location from several trained locations while walking. Kunze et al. (2005) assume that the sensor's location discovered while the subject was walking would carry onto the next activity.

One challenge of having a sensor at an unknown orientation is that deducing the subject's posture is difficult if not impossible since deducing a particular posture depends on knowing the orientation of a particular body-location. Because of that Quwaider and Biswas (2008) propose using the RSSI of the communication between sensing nodes placed on various parts of the body to detect posture. Variations in clothing and subject behavioural differences cause them to use a HMM instead of a simpler threshold-based method.

As an alternative Bieber et al. (2009) grouped their standing, sitting and lying down activities as "resting" thereby avoiding classifying each of the postures separately. Such "resting" postures can be easily differentiated from more active activities by analysing the frequency distribution. Quwaider and Biswas (2008) noted a lack of frequency components below 2 Hz. Perhaps a more accurate cut off frequency would be 1.35 Hz, which is the lowest step frequency found by Oberg, Karsznia, and Oberg (1993). The frequency was the 95% confidence level lower bound for a slow gait of females within the age groups 10-14 and 60-69.

To summarise, there are several ways of dealing with the variety of phone orientations in the literature:

1. Use orientation-invariant data (such as the magnitude of the acceleration vector) to compute features.
2. Change the orientation of the data to world coordinates before computing features. To do this, one can use the gravity component from the accelerometer to compute where down is relative to the phone, then extract features from the sensors values relative to where down is. This approach has been shown to perform better than the approach given in item 1 (Yang, 2009).

   This approach can further be broken down into:

   a) After changing data to world coordinates, two sets of features to be extracted: those in the direction of the gravity vector, and those in the plane perpendicular to the gravity vector (the horizontal plane). Only the magnitude of values in the horizontal plane is taken because of lack of direction.
   b) From data extracted from the horizontal plane as obtained in 2a, a general direction of the accelerations is taken to be one of the two orthogonal

axi in the horizontal plane. This is performed by using PCA. By so doing, 3 axi are obtained: up-down axis parallel to gravity, the forward-backward axis parallel to the main direction of motion which lies on the horizontal plane, and the sideways axis which lies on the horizontal plane but is perpendicular to the forward-backward axis. This method has been shown to achieve slightly better results than 2a, but faces the same issue of rotation explained in 2a.

2.5 Classification component

This section will give research conducted in areas related to the classification component of an activity recognition system (Choudhury et al., 2008).

Comparison of classifiers

Korel and Koo (2007) compared Artificial Neural Networks (ANNs), Kohonen Self-Organising Maps (KSOMs), Bayesian Networks (BNs) and Hidden Markov Models (HMMs) and concluded that none is the 'best' for deducing the context but each addresses different issues that arise from recognising context in Body Sensor Networks (BSNs).

Some of the challenges facing context recognition in BSNs outlined by Korel and Koo include: the presence of noise; the continuity of human movement compared to the instantaneous classification of context; new contexts may be continuously added to the system of the life-time of the system; dealing with high data dimensionality due to obtaining input from multiple sensors mounted in multiple body-locations on the user's body; and recognising which inputs (from all the sensors available to the system) are relevant to the identification of the user's context.

Korel and Koo explain some of the key advantages of using ANNs is their ability to deal well with noise from sensors and the possibility of unsupervised training.

This allows the system to undergo training while in use. They also explain that KSOMs are advantageous in that they require no prior information on the contexts that the system would encounter while in use. Hence they are able to deal well with data that is not labelled or data from contexts that are not previously defined or those that are unpredictable. However, Korel and Koo explain that KSOMs suffer from the *curse of dimensionality*. As more inputs become available the KSOM tends to grow slow and less fault tolerant since it uses more resources to map the high dimension input space to a large output space.

For BN, Korel and Koo explain that the graphical probabilistic models result in high accuracy. However, confusion in the activity labelling or insufficient sample sizes in the training data results in lower activity classification accuracy. Korel and Koo further explain that previous work that used BNs showed high accuracy rates in controlled environments but suffered in uncontrolled environments. This is possibly due to the assumption that all attributes that influence a classification decision are observable. They add that BNs are limited by their inability to exhaustively model all relationships in an uncontrolled environment.

For HMMs, Korel and Koo explain that due to the probabilistic foundations allow HMMs to deal well with sensor noise. In addition, HMMs are able to deal with time variations, repetitions and variable length sequences. However, HMMs require a training phase that optimally should occur without user intervention. They note that some instances of HMMs have been proposed in literature that can deal with online training.

Supervised vs unsupervised training & testing data

Ermes and Juha (2008) analysed the use of data gathered in a supervised and unsupervised environment to train and test activity recognition systems. They found that although training and testing with unsupervised data achieved a high success rate of 89%, this success rate falls dramatically to 72% when supervised data is used for training and unsupervised data for testing.

Need of specific user training

Saponas, Lester, Froehlich, Fogarty, and Landay (2008) analysed whether it is necessary for a user to train an activity recognition system using his or her own motions. For the 8 subjects studied and 4 activities (walking, running, cycling and sitting), they found high success rates for both models trained and tested from the same subject (mean 99.48%, s.d. 0.91%) and models trained by one subject and tested with another (mean 97.4%, s.d. 4.05%). They concluded that it was not necessary for a user to train an activity classification system with his or her own personal data.

Continuous activity classification

Based on the lessons learnt in their project, Choudhury et al. (2008) explain that activity classification from features extracted from a single point of time fails to take into account the temporal continuity of activities and leads to "choppy classification" where a single activity may be broken down into several segments. This problem particularly affects applications that take actions depending on the activity recognised.

User perception of inference errors

Consolvo et al. (2008) explain that while traditional classification results are grouped into four (true positive, false positive, true negative, false negative), the error metrics fail to "capture subtleties of how users perceive inference errors". He further outlines seven types of errors perceived by users of their fitness device:

1. Activity start time error.

2. Activity duration error.
3. Confuse an activity the system **was** trained to infer with another activity it **was** trained to infer.
4. Confuse an activity the system **was not** trained to infer with an activity it **was** trained to infer.
5. Fail to detect an activity the system **was** trained to infer.
6. Fail to detect an activity the system **was not** trained to infer.
7. Detect an activity when none occurred.

Consolvo et al. (2008) explain that within these errors, some errors were found to be intolerable while others were tolerable by system users. Among the errors, 5 caused the users to became frustrated because they had gone to the trouble of wearing or using the system while performing the physical activity and 7 decreased the credibility of the system. 6 was not considered an error by the users but some users were disappointed because they had hoped that the system would at least register some physical activity. Users did not mind 3 since they could easily change it to the correct activity as long as the start time and duration were correct, some users appreciated 4 when they did an activity the device was not trained for and the device registered a physical activity, while 2 only mattered to the users when the registered duration meant them not achieving their desired goals.

Simple vs complex activities

Kim et al. (2010) explain that there are currently challenges that are faced with recognising complex activities. Some of these challenges pertaining to activity recognition using wearable sensors are:

1. *Concurrent activities*

   People can undertake multiple activities simultaneously. An example is watching TV while talking. Current activity recognition has focussed on recognising sequential activities. Recognising such concurrent activities requires a different approach than what is currently taken in activity recognition research.

2. *Interleaved activities*

   Certain activities can be interleaved. An example is a person could be cooking while doing other activities. From time to time, the person checks on the status of the cooking, perhaps does something related to the cooking, then returns to the other activity.

3. *Interpretation ambiguity*

   Depending on the context, certain actions could be interpreted as different activities. An example being a person opening the fridge. This could be interpreted as cooking in one circumstance or cleaning the fridge in another.

Dernbach et al. (2012) attempted complex activity recognition using a smart phone. Using a single carried smart phone, they achieved a high success rate classifying simple activities but did not achieve a good success rate classifying complex activities. The simple activities include: biking, climbing stairs, driving, lying, running, sitting, standing and walking. The complex activities include: cleaning (wiping down a kitchen counter top and sink), cooking (heated a bowl of water in a microwave and poured a glass of water from a pitcher), medication (subject retrieved pills from a cupboard and sorted out a week's worth of doses), sweeping (sweeping the kitchen area), washing hands (the subject washed hands using soap at the kitchen sink), watering plants (the subject filled a watering can and watered three plants in two rooms).

It should be noted that Dernbach et al. (2012) applied similar methods for classifying both simple and complex activities. Their method involves gathering samples from the accelerometer and magnetometer on the phone, grouping the samples into windows, extracting statistical features (mean, minimum, maximum, standard deviation, zero-crossing counts and correlation) from the windows, and classifying using various classifiers. The issues outlined as challenges of complex activity recognition (concurrent and interleaved activities and interpretation ambiguity) are not dealt within their algorithms. Dernbach et al. (2012) explain that some of the challenges of complex activity recognition can be solved by having more sensors placed in the right locations (in the environment) to recognise the activity.

2.6 Comparison of cross-validation techniques

Two primary cross-validation techniques exist in the activity recognition literature: N-fold and remove-one-subject. For N-fold cross-validation the data is divided into N equal folds. In each test iteration, one randomly selected fold is used for testing and the rest are used for training. Hence, all subjects contribute to the training phase. For remove-one-subject cross-validation the data is divided into groups, each group representing one subject. In each test iteration, one randomly selected subject is then used for testing while the rest are used for training.

Lockhart and Weiss (2014) categorise the models used in activity recognition into three:

1. Impersonal models: the system is trained by a number of subject but then used by a new subject.
2. Personal models: the system is only trained by the user who will utilise it.
3. Hybrid models: the user who ends up utilising it is part of the set of users who trained the system.

It should be noted that it is possible for a system to make use of different models at different times. For example, a system could come preloaded with an impersonal model, then by incorporating the data of it's current user into it's dataset, shift to a hybrid model and perhaps eventually end up with a personal model of that user.

Recent literature (Lockhart & Weiss, 2014; Patel et al., 2009; Reiss, 2014) has shown that impersonal models have lower success rates and higher variance in their testing results while the results of personal or hybrid models tend to be "optimistic".

The lower success rates achieved by impersonal models is due to inter-subject variability. For example, walking patterns of a young person could be different from those of an elderly person. This suggests that the best success-rates of an end system are achieved by having the end system adapt to the end user. In addition, it also suggests that validation results from personal models and hybrid models represent the best-case-scenario achievable by activity recognition systems.

2.7 Other research

Battery consumption

Battery consumption is an important issue for wearable systems including activity recognition systems that use wearable sensors. The battery time limits the length of time that the system can be used before a recharge. Larger batteries make systems less portable and more cumbersome. However, accurate activity recognition at many times involves heavy data processing which consumes more battery power. Hence, in many cases, a balance has to be maintained between data processing (hence more accurate activity recognition) and battery life. Better algorithms that allow more accurate activity recognition with less data processing are necessary.

In their work, Wang et al. (2009) show that they could increase the phone's battery life by more than 75% while at the same time maintaining high accuracy and low latency in identifying user states by selecting the most appropriate sensor to use from a large set of sensors and when to use the sensors. The user state included the activity, location, amount of motion and type or amount of background sounds.

Classifying atomic body movements rather than activities

Physical activities can be thought of as being composed of smaller motions. These smaller motions (atomic motions), when combined form simple activities (when only a few atomic motions are present) and complex activities (when many atomic motions are present) (Dernbach et al., 2012).

Ghasemzadeh et al. (2008) propose a linguistic framework for encoding activities. The framework involves encoding the activities under classification as primitives which represent atomic body movements. These primitives form the building blocks of activities. This approach mirrors a similar approach taken in speech recognition where a word is broken down to phonemes. Classification occurs first at the phoneme level, the results of which are combined to the most likely word.

In their paper, Ghasemzadeh et al. (2008) used 18 sensors (each with a 3D accelerometer and a 2D gyroscope) spread across the body to record inertial readings of a selection of atomic body movements. The recordings were then manually segmented and a number of features extracted from them. The features were then used to formulate the representations of the atomic activities observed from each sensor location. During testing each sensor location classifies its readings to the most likely primitive. The aggregate of the primitives from all the sensor locations can then be aggregated to determine the current activity.

A general weakness with the approach used by Ghasemzadeh et al. (2008) is that while it is relatively simple to manually deduce, record and segment the atomic activities of repetitive activities like walking and running, it is much more difficult (and potentially impossible) to do so for more complex activities like exercise routines and household chores. This is especially so if the data gathering is performed in an uncontrolled environment where labelling of atomic activities within a larger activity presents both an inconvenience to the subject and introduces errors into the data whenever the subject stops to mark the end of an atomic segment.

In such cases it is necessary to automate the identification of atomic activities. Additional advantages of automated identification of atomic activities as noted by Minnen, Starner, Essa, and Isbell (2006) includes:

1. Reducing the cost of data gathering by reducing the amount of labelling re-

quired.

2. As an exploratory tool to ease the burden while analysing large amounts of activity data.
3. Validation or alternatives to atomic activities for activities we believe we understand.
4. Possibility of systems that automatically adapt to new body sensor locations and activities.

2.8 Current state of commercial on-body activity trackers

At the time of writing, systems that are wearable, not specific to particular activities and monitor the user's activity are in the form of activity trackers. These are devices that monitor an individual's fitness level by monitoring his/her level of activity during the day and include popular devices like FitBit [1], Nike+ FuelBand SE [2], Garmin Forerunner 910XT [3], Garmin Vivofit [4], etc. Devices exist that monitor measurements of a specific activity (e.g. running, cycling, swimming, etc.) but those are not considered here.

Beckham (2012) explain that these devices are aimed towards combatting obesity and the main reason why they are effective is psychological: by measuring activity levels, they provide data that can be used to set goals that when achieved provide a sense of accomplishment which encourages the user to set more goals. Activity trackers provide a way to objectively measure a person's activity levels, something that people are not good at doing. In addition, not only have these systems gamified

[1] http://www.fitbit.com/au

[2] http://www.nike.com/us/en_us/c/nikeplus-fuelband

[3] http://sites.garmin.com/en-AU/forerunner910xt/

[4] http://sites.garmin.com/en-US/vivo/vivofit/

exercise making it fun, but the ability to post results on social media allows the users to boast about their achievements, possibly be encouraged by his/her peers to keep up his/her achievements and form a community of enthusiasts that help keep each other fit.

According to Reviews (2014), some of the features common to activity monitors include:

1. Step counting: almost all devices count steps.
2. Calories burnt estimation: almost all devices offer an estimation of calories burnt.
3. Capability to synchronise captured data: currently many activity trackers offer wireless synchronisation while a few offer wired synchronisation.
4. Data visualisation: many devices offer visualisation either on the device, on a mobile device, on the web or a combination of these.
5. Differentiating stairs from steps on flat ground: only a handful of devices are able to differentiate steps performed on stairs from flat ground.
6. Form: many forms exist ranging from: adhesive patch, ankle band, arm band, wrist band, watch, clip, key chain, pendant, headband, headphones, lower-back strap, mounted pod, ring, A few devices can be used in a number of forms.
7. Heart rate monitoring: some systems offer heart rate monitoring in addition to other activity level measurements. Some systems do this on the same device while some require additional devices.
8. Sleep monitoring: some systems monitor the subject's sleep in addition to activity levels while awake.
9. Food logging: some systems offer a way, either through a mobile device or on the web, for the user to log his or her food intake.

As noted above, almost all commercial activity trackers count steps but few differentiate between different kinds of steps (e.g. walking on flat ground versus walking up stairs). This agrees with Consolvo et al. (2008) when they note that most of the pedometers of the day have a simple "inference model" that interprets a step as a sequence of ascending and descending accelerations. This makes these devices inaccurate for measuring the daily energy levels that they set out to measure. Better recognition of the user's activity would improve this measure.

2.9 Conclusion

To conclude the literature review, the following are the research questions that appear in the thesis and their links back to the activity recognition using on-body inertial sensors literature.

Given the data available from current IMUs - accelerations, velocities, and orientations - which individually and in combination will provide the best activity classification performance in terms of success rate and other settings?

The data that is readily available from an IMU (without integrating or differentiating) are: accelerations (sensor readings from an accelerometer), rotational velocities (sensor readings from a gyroscope) and orientations (the combination of accelerations, rotational velocities and readings from a magnetometer normally performed within the IMU resulting in the orientation of the device relative to gravity and the magnetic north).

From these three *sources*, the next sections will propose research questions aimed to select the *source* that is best used for activity recognition based on the success-rate that can be achieved and the sampling requirements of the *source*.

Which of the three *sources* requires the lowest sampling frequencies to achieve the highest activity classification success-rates? Which one requires the highest?

As explained in section 2.3, several sampling frequencies have been used in literature. However, little analysis has been performed on the impact of sampling frequencies on activity classification success-rates using on-body inertial sensors. The two analyses found are: Bieber et al. (2009) who explain that a sampling frequency of 32Hz is sufficient to sample body movements based on the response time of human reflexes, and Maurer et al. (2006) who analysed the obtained activity classification success-rates for sampling frequencies in the range [0Hz,50Hz]. Maurer et al. only analysed success-rates obtained from feature-vectors extracted from accelerations. There has to date, been no communication in the research literature, concerning the impact of sampling frequency upon activity classification accuracy while using feature-vectors extracted from rotational velocities or orientations.

In addition, their study only included stationary activities (like sitting and standing) and ambulatory activities (walking on a flat surface, walking up stairs, walking down stairs and running). However, it is likely that other Activities of Daily Living have different sampling requirements to those observed by Maurer et al.

Hence, section 4.2 analyses the differences in the sampling requirements of the three *sources* with relation to the accuracy of recognising Activities of Daily Living, using multiple on-body inertial sensors that have fixed orientations relative to the body-locations mounted on and fixed locations on the subject's body. This is performed in order to determine which *source* requires the highest sampling frequency and which *source* lowest sampling frequency in order to achieve the highest activity classification success-rates.

How do the success-rates obtained from data of each *source* compare those from other *sources* and to the three *sources* combined?

As explained in section 2.3, activity classification systems have been proposed in the literature that make use of multiple types of inertial sensors. These include systems proposed by Amft and Tröster (2008); Kunze and Lukowicz (2008); Yuwono et al. (2012) and Dernbach et al. (2012). However, little analysis has been performed to ascertain the capabilities of different on-body inertial sensor types in recognising Activities of Daily Living. The closest work found that performed this is Dernbach et al. (2012) and Shoaib et al. (2014).

Dernbach et al. compared the success-rates achieved from feature-vectors extracted from accelerations and those achieved from feature-vectors extracted from accelerations and orientations for a smart-phone-based activity recognition system. They concluded that using orientations in addition to accelerations resulted in 10-12% increase in activity classification success-rates than was obtained by using accelerations only. Dernbach et al.'s data was gathered using a smart-phone that was carried in different orientations and body-locations from one subject to the next. It is likely that different results would be observed in a case where multiple sensors are used per subject, and the location and orientation of the sensors relative to the body-location are fixed (as is the case with wearable-sensor-based activity recognition).

Shoaib et al. compared the success-rates achieved by feature-vectors extracted from accelerations, rotational velocities and compass direction, in recognising stationary (sitting and standing), ambulatory activities (walking on a flat surface, walking up stairs, walking down stairs, jogging, running) and biking. Although data of multiple body-locations was gathered, the data was not combined to take into account readings observed at different body-locations within each window. Hence, like Dernbach et al.'s analysis, Shoaib et al.'s analysis is focussed toward smarthone-based activity recognition. They concluded that, with the exception of the magnetometer, that accelerations and rotational velocities have the potential to individually result in high activity classification success-rates, depending on the activity and body-location. They also conclude that the sensors need only be combined in the case when both sensors are not performing well.

Hence, section 5.4 compares the activity classification success-rates obtained from feature-vectors extracted from each individual *source*. In addition, a comparison is performed between activity classification success-rates obtained from feature-vectors extracted from each individual *source* to those obtained from feature-vectors extracted from all three *sources* combined. The analysis is performed within the context of wearable-sensors and hence makes use of data captured from multiple on-body inertial sensors that have fixed orientations relative to the body-locations mounted on and fixed locations on the subject's body.

What is the relationship between activity classification success-rates and the length of the windows used to extract features for classification?

As explained in section 2.4, several window lengths have been used in the activity classification literature. However, little analysis has been performed on the impact of window lengths on activity classification success-rates using on-body inertial sensors. The analyses found are: Huynh and Schiele (2005); Patel et al. (2009) and Dernbach et al. (2012).

Huynh et al. analysed window lengths in the range [0.25s,4s] on ambulatory activities and riding the bus and concluded that no single window length is best for all activities. The one second window performed best when the activity being tested was walking or jogging, the 2 and 4 second windows performed best for skipping and hopping, while the 0.25 and 0.5 second windows performed best for standing.

Dernbach et al. analysed the impact of window lengths in smart-phone-based activity classification success-rates for both simple and complex activities. Dernbach et al. found that shorter window lengths performed significantly better than longer window lengths. However, Dernbach et al. used a single sensor per subject which was carried in the body-location and orientation that the subject prefered. It is possible that the variation in sensor orientation and body-location have an impact in the classification results obtained. In which case, the relationship might not hold in a case where multiple sensors are used per subject, and the location and orientation of the sensors relative to the body-location are fixed (as is the case with wearable-sensor-based activity recognition).

Contrary to Dernbach et al.'s finding, Patel et al. found that smaller window lengths resulted in lower activity classification success-rates and increasing the window length past 6 seconds only results in marginal increases in activity classification accuracy. However, Patel et al.'s research is primarily focussed on recognising physical activities of subjects suffering from Chronic Obstructive Pulmonary Disease (OCPD). Hence, all their subjects that were suffering from OCPD.

Hence, section 4.3 analyses the relationship between the length of the windows used to extract features and the accuracy of recognising Activities of Daily Living in able-bodied subjects, using multiple on-body inertial sensors that have fixed orientations relative to the body-locations mounted on and fixed locations on the subject's body.

What is the relationship between activity classification success-rates and overlap size of the windows used to extract features for classification?

As explained in section 2.4, several window overlaps have been used in literature. However, there has, to date, been no communication in the research literature, concerning the impact of window overlaps upon an activity classification algorithm's accuracy.

Hence, section 4.4 analyses the relationship between the amount of overlap in the windows used to extract features and the accuracy of recognising Activities of Daily Living, using multiple on-body inertial sensors that have fixed orientations relative to the body-locations mounted on and fixed locations on the subject's body.

Is there a significant difference in activity classification rates when a monitor is mounted on different body-locations? If so, which body-location yields the highest activity classification accuracy?

Several body-locations on subjects' bodies have had monitors mounted on them in the literature. However, little analysis has been performed to ascertain the relative merits of different body-locations in recognising Activities of Daily Living. As explained in section 2.3, the closest work found that performed this is Henpraserttae et al. (2011); Keally et al. (2011); Maurer et al. (2006) and Shoaib et al. (2014).

Keally et al. analysed the success-rates achieved by accelerometers mounted on different locations on the body in recognising postures (sitting and standing), ambulatory activities (walking on a flat surface, walking up stairs and walking down stairs), writing on a whiteboard, typing on a keyboard and shaking hands. They concluded that the postures and ambulatory activities could be identified equally well with any sensors on the leg. Data gathered from accelerometers mounted on the upper body resulted in higher success-rates for writing on a whiteboard, typing on a keyboard and shaking hands. Within these three activities, data gathered from sensors mounted on the right side of the upper body resulted in higher success-rates for shaking hands while data gathered from sensors mounted on the right side of the upper body resulted in higher success-rates for writing on a whiteboard and typing on a keyboard. However, Keally et al.'s data was gathered from one subject only.

Maurer et al.'s research included stationary activities (sitting and standing) and ambulatory activities (walking on a flat surface, walking up stairs, walking down stairs and running) for six subjects. The body-locations on which monitors were mounted are: left wrist, belt, necklace, right trouser pocket, shirt pocket and bag. Maurer et al. noted that features derived from a 2D accelerometer mounted on the subjects' bag resulted in higher success-rates than those derived from accelerometers mounted on the left wrist, belt, necklace, right trouser pocket and shirt trouser pocket.

Henpraserttae et al's research primarily analysed the impact of accelerometer signal transformations from local to global coordinates on smart-phone-based activity classification. However, part of their work analysed the performance obtained from data captured at different body-locations. The locations tested only included locations on which a phone could be carried: shirt-pocket, trouser-pocket and waist. They found that the waist results in the highest success-rate, followed by the shirt-pocket and trouser-pocket.

Although Shoaib et al.'s research primarily analysed the activity classification success-rates obtained by sensors available in smart-phones (accelerometers, gyroscopes and magnetometers), their work also included an analysis of activity classification success-rates of several locations on the body where phones are carried. The locations analysed are: left trouser pocket, right trouser pocket, belt position towards the right leg, upper right arm, lower wrist. Shoaib et al. conclude that the location that results in the best activity classification success-rates varies from activity to activity.

Hence, section 5.4 compares the success-rates of recognising Activities of Daily Living obtained from six body-locations using data gathered from 20 subjects. The body-locations include: the dominant upper arm, dominant wrist and non-dominant wrist on the arms; the chest; and the thigh and ankle on the right leg. The monitors have a fixed orientations relative to the body-locations mounted on and features extracted from accelerations, rotational velocities and orientations are combined for each extracted window.

What is the functional relationship between activity classification success-rates and the number of monitors mounted at different locations on the body?

As explained in section 2.3, the closest research in the literature that analysed the relationship between the number of sensors mounted on the subject's body and the activity classification success-rates obtained is that of Zappi et al. (2007).

Zappi et al. compared the number of monitors used to recognise activities of automotive assembly line workers and found that a monitor mounted on one body-location could at best yield a success rate of 50% rising to 80% for 3 body-locations and 98% for 57 body-locations. This indicates that there is a depreciating returns relationship between the number of body-locations monitored and the accuracy of recognising activities in automotive assembly line work. Similar work for Activities of Daily Living was not found.

Hence, section 5.6 analyses the relationship between the number of body-locations monitored and the success-rates of recognising Activities of Daily Living obtained. Up to six body-locations are analysed that include: the dominant upper arm, dominant wrist and non-dominant wrist on the arms; the chest; and the thigh and ankle on the right leg. The monitors have a fixed orientations relative to the body-locations mounted on and features extracted from accelerations, rotational velocities and orientations are combined for each extracted window.

Are some activities more easily identifiable than others? Which activities are confused with one another?

As explained in section 2.5, Dernbach et al. (2012) noted that activities like biking, climbing stairs, driving, lying, running, sitting, standing and walking, can be identified more accurately than activities like cleaning (wiping down a kitchen counter top and sink), cooking (heated a bowl of water in a microwave and poured a glass of water from a pitcher), medication (subject retrieved pills from a cupboard and sorted out a week's worth of doses), sweeping (sweeping the kitchen area), washing hands (the subject washed hands using soap at the kitchen sink), watering plants (the subject filled a watering can and watered three plants in two rooms). Other publications (such as Bao and Intille (2004); Ermes and Juha (2008)) also noted differences in success-rates between activities but did not notice any grouping of activities that perform better than others.

Based on the user perception of activity classification errors explained in section 2.5, Taylor et al. (2011) explain that activity classification systems that recognise Activities of Daily Living should select more specific activities (e.g. dicing, grating or peeling) only if the certainty of the more specific activities is very high, otherwise the activity classification systems should favour more generic activities (e.g. cooking). Hence, an analysis comparing the mutual confusion rates of specific activities needs to be performed. This analysis could assist in selecting when a more specific activity should be picked instead of a more generic activity. There has, to date, been no communication in the research literature, concerning the analysis of mutual confusion between Activities of Daily Living.

Hence, section 6.2 compares the accuracy at which different activities within Activities of Daily Living are identified while section 6.3 analyses the mutual confusion between different activities. The analysis is performed with accelerations, rotational velocities and orientations from six body-locations are combined for each extracted window. The activities studied include ambulatory activities (walking on a flat surface, running, walking up and down stairs), desktop activities (writing and using a PC), communication activities (texting and talking on the phone), laundry activities (ironing and folding clothes), household cleaning (sweeping, vacuuming and dusting), food preparation activities (dicing, grating, peeling, stiring, washing dishes and washing hands) and other activities (watching TV and brushing teeth).

How different are activity classification success-rates obtained from 10-fold cross-validation from those of remove-one-subject cross-validation?

As explained in section 2.6, recent literature by Lockhart and Weiss (2014); Patel et al. (2009) and (Reiss, 2014) has shown that impersonal models (i.e. those where the training set and testing set contain different subjects) achieve lower success rates and higher variance in their testing results than the results of personal (i.e. the testing set and training set contain data from the same subject) or hybrid models (i.e. some of the subjects whose data is in the training set are also in the testing set).

However, it is still unclear how different results of 10-fold cross-validation would be from those of remove-one-subject cross-validation or whether any relationship between the two result sets exist. Are relationships observed using 10-fold cross-validation valid for remove-one-subject cross-validation? Hence, section 6.4 compares the activity classification success-rates obtained using 10-fold cross-validation to those obtained using remove-one-subject cross-validation. An important research question answered is whether a correlation exists between result sets obtained from 10-fold cross-validation and those obtained using remove-one-subject cross-validation. The existence of a correlation implies that certain relationships observed in 10-fold cross-validation results might also be valid for remove-one-subject cross-validation.

Can the body-locations at which a sensor is mounted on a user be identified without knowing the activity that the user is undertaking?

As explained in section 2.3, phones are carried on several different locations on the body. In addition, the carry location varies from person to person and from occasion to occasion. In addition, as noted by Henpraserttae et al. (2011), motion patterns captured at different body-locations are significantly different from each other, such that training a classifier with data from one body-location results in low activity classification success-rates when tested with data from another body-location.

As explained in section 2.4, among the proposed solutions in literature include training a meta-classifiers with one classifier per body-location (Sun et al., 2010) and deducing the body-location by classifying patterns observed at different body-locations in a particular activity like walking (Kunze et al., 2005). The first approach implies multiple classifiers need to run on the activity classification system hence taking up more resources. The second approach does not take into consideration any changes in the location on which the monitor is mounted that might occur between occurences of the selected activity. Even though the walking activity, as selected by Kunze et al., occurs frequently in the day-to-day life of users, it is likely that the user could change the location of the monitor in between walking periods.

Hence, section 7.1 analyses the ability to identify the body-location on which a monitor is mounted on while the user is performing an Activities of Daily Living using a single classifier (without using a meta-classifier) but without the knowledge of which activity the user is currently carrying out. The analysis is performed with smart-phone-based activity recognition in mind. The ability to identify the body-location on which a phone is carried, could allow smart-phone-based activity classification to take advantage of that information to allow for more accurate activity classification.

What is the impact of random monitor rotations on activity classification success rates?

One of the biggest challenges facing smart-phone-based activity classification is that the phone can be carried in any orientation relative to the body-location it is monitoring. Three possible solutions have been proposed in literature: using features that are independent of the orientation of the phone (such as the magnitude of the acceleration vector); using a meta-classifier that has one classifier per possible device orientations (Sun et al., 2010); and reorienting the sensor signal from the sensor's local coordinates to global coordinates (such that the resultant coordinate axes are aligned to gravity and either the magnetic north or the subject's direction of motion).

The first method has been shown to result in lower success-rates than the third method (Yang, 2009). The second method has been noted to be impractical since many device orientations are possible (Henpraserttae et al., 2011; Yang, 2009). The third method has been shown to result in high activity classification success-rates (Henpraserttae et al., 2011; Yang, 2009).

The method of reorienting the sensor data from local to global coordinates was initially proposed by Mizell (2003). Mizell proposed computing the mean of the accelerations within a window (i.e. the gravity component) and using it to reorient the sensor data. This serves as a low pass filter that removes any accelerations due to body motions. The gravity component can then be used to extract the sensor values that are in the vertical world axis. The method was later used by Yang together with the magnitude of the horizontal components. Henpraserttae et al. later refined the method further by extracting the anteroposterior (forward-backward) axis from the horizontal components by using the PCA of the horizontal components of the accelerations.

However, the method of extracting the gravity component proposed by Mizell (and later used by Yang and Henpraserttae et al.) suffers when the accelerometer changes orientations relative to gravity. This is bound to happen whenever the subject moves his or her limbs and results in a gravity vector that lies between the initial sensor orientation and the final sensor orientation. This then results in errors in the reoriented signal values.

IMUs, however, compute the orientation of the sensor by integrating the rotational velocities obtained from the gyroscope then including the gravity vector computed from the accelerations to correct the eventual drift. Hence, the orientation of the monitor relative to world coordinates at each sample is accurately known even when the orientation is changing.

Hence, section 7.1 analyses the impact random sensor rotations have on activity classification success-rates and how well the two reorientation methods solve the problem. First, the activity classification success-rates obtained from data with and without random rotations are compared. Next, the success-rates obtained from data reoriented using the accelerometer-based method (as proposed by (Henpraserttae et al., 2011)) are compared to when reorientation is not performed and to when the orientation-based method (using the orientation data provided by an IMU) is performed.

3 Methodology

3.1 Sensor description

An Opal System from APDM (APDM, 2012) was used in wireless buffering mode to capture data. Each sensor module (monitor) contains a triaxial accelerometer, magnetometer and gyroscope. The properties of the sensors are shown in table 3.1. Sampling was performed at 128Hz and all the modules were synchronised ($\leq 10\mu s$ synchronised sample timing difference).

Simoes (2011) make a comparison of gait parameters (cadence, torso rate of rotation, head rate of rotation and stride length) extracted using these sensors against a camera-based Vicon system (referred to as the 'gold standard' in kinematic motion tracking by Simoes (2011)). High correlations were found between the gait

Table 3.1: Properties of the sensors

Property	Accelerometer	Magnetometer	Gyroscope
Axes	3	3	3
Range	±6g	±6 Gauss	$2000deg/s$
Noise Density	$0.0012\ m/s^2/\sqrt{Hz}$	$0.5mGauss/\sqrt{Hz}$	$0.05deg/s/\sqrt{Hz}$
Sampling Rate	1280Hz	1280Hz	1280Hz
Output Rate	128Hz	128Hz	128Hz
Bandwidth	50Hz	50Hz	50Hz
Resolution	14bits	14bits	14bits

Table 3.2: Sensor placement locations.

Location	Description
Ankle	Just above the ankle joint facing outward.
Thigh	In the thigh pocket, facing forward.
Chest	On the chest, above the sternum, facing forward.
Dominant upper hand	Between the shoulder and the elbow, facing outward.
Dominant wrist	Approx. 3cm from the wrist-joint, lying flat against the back of the dominant arm, facing outward.
Non-dominant wrist	Approx. 3cm from the wrist-joint, lying flat against the back of the dominant arm, facing outward.

parameters extracted by the two systems.

Although part of the potential application of this research is the use of orientation data from mobile devices, we did not use an actual phone for data capture to avoid inconsistent sampling rates. Sampling rates vary from phone to phone and depend on the load on the phone at the time of sampling (Dernbach et al., 2012; Taylor et al., 2011).

An image of one of the monitors is given in figure 3.1(a). Figure 3.1(a) shows a monitor attached to wrist/ankle straps. The monitors can be attached to a variety of straps. Small straps that can be used to strap the monitor onto the wrists and ankle are shown in figure 3.1(c). Medium sized straps that can be used to strap the monitors onto the upper arm are shown in figure 3.1(b). Larger staps that can be worn around the chest to secure the monitor onto the sternum are shown in figure 3.1(d).

3.2 Sensor placement

The monitors were placed in six body-locations on each subject. The body-locations are described in table 3.2 and illustrated in figure 3.2(a). An example of a subject wearing the monitors is shown in figure 3.2(b).

(a) Monitor attached to wrist/ankle strap

(b) Upper arm Strap

(c) Wrist/Ankle Strap

(d) Monitor attached to sternum Strap

Figure 3.1: Monitor and straps

(a) Sensor placement model (b) Sensor placement on a subject

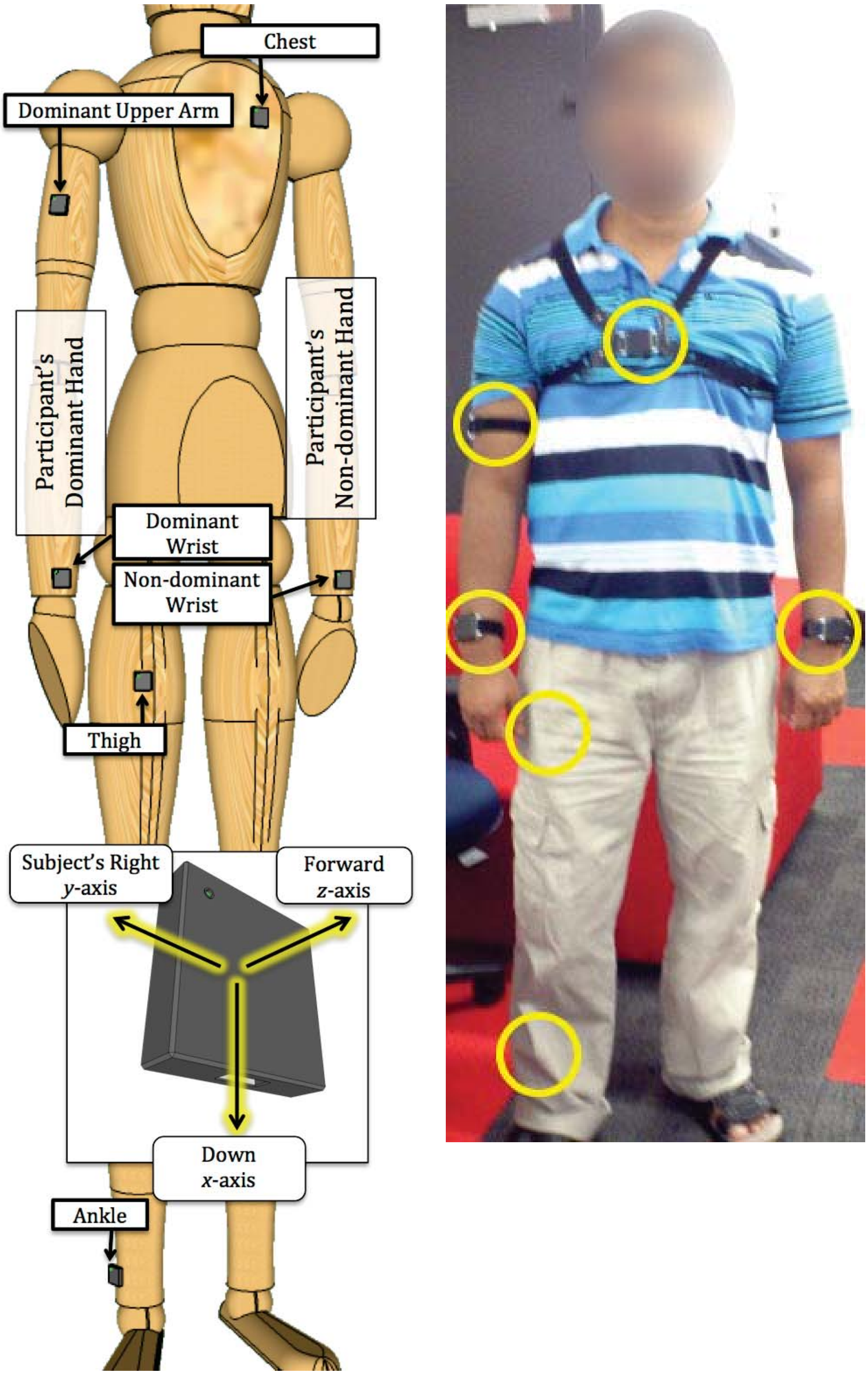


Figure 3.2: Sensor placement model and placement on a subject

(a) Length

(b) Width

(c) Thickness

Figure 3.3: Thigh monitor case

For walking and running, the monitors were placed in three of these six body-locations. These were: the dominant wrist, the chest and the thigh location. More details on why this was done are given in section 3.3.

For each body-location, the monitor was placed such that x-axis faced downward for all body-locations while the subject stood upright with arms in a downward rested position. The y-axis and z-axis of the monitor were both in the horizontal plane, with the z-axis facing perpendicular to the surface of the body-location the monitor was strapped on. Where possible the monitor was placed such that the z-axis faced forward (e.g. chest and thigh). In other body-locations the z-axis either

Figure 3.4: Inside the thigh monitor case

faced sideways (ankle) or 45°between sideways and forward (wrists).

Strapping a monitor around the thigh was noted to be uncomfortable for the subject while gathering walking and running data. To avoid this, the monitor was placed in a larger case (resembling the size of a mobile phone) and placed upright in the subject's pocket so as to be in the orientation shown in figure 3.2(a) for all other activities. The subjects were asked to empty the pocket so as to avoid any interference between the sensor and any other objects that might be in the pocket (e.g. keys, phones). None of the subjects had thigh pockets that were large enough for the thigh monitor (while in the case) to move around freely within the pocket. The case is shown in figure 3.3. The dimensions of the case were 12cm by 6cm by 2cm and are shown in figure 3.3(a), figure 3.3(b) and figure 3.3(c) respectively.

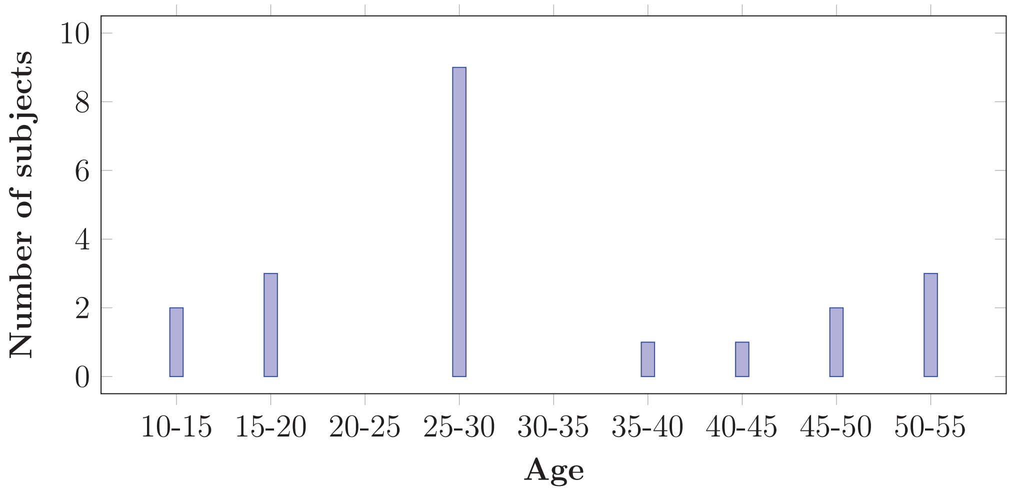


Figure 3.5: Distribution of subject ages for the first data gathering session (walking and running).

Figure 3.4 shows the inside of the phone case. The monitor was placed on the upper part of the case, and secured in place using velcro.

All other monitors, except for the thigh monitor, were secured so as not to move freely relative to the body-location they were placed on. However, the straps of the monitors were loose enough to be comfortable for the subject.

3.3 Data gathering

The data gathering was divided into two sessions. The two sessions are separated in time by two years and hence include different subjects.

In the first session only walking and running data was gathered. The data was gathered outdoors upon a flat soccer field. Only three monitors were available to gather the data with during the first session. The data of 21 subjects was gathered. The subjects included 14 males and 7 females, with ages ranging from 14 to 51.

In the second session, an additional 20 activities of daily living were gathered from 18 subjects. Some of the activities can be categorised together to constitute some day-to-day routines. For instance, vacuuming, cleaning and dusting together constitute the routine of household cleaning. The activities are as follows (more

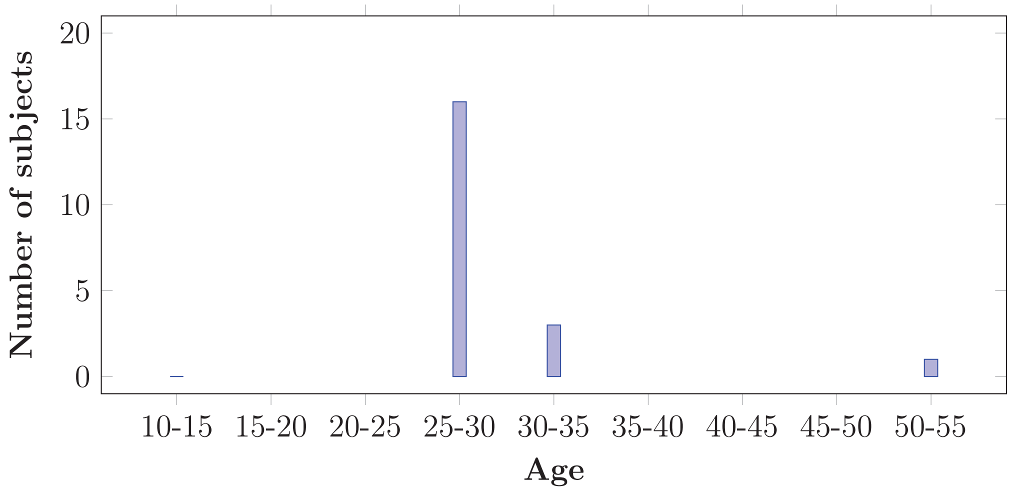


Figure 3.6: Distribution of subject ages for the second activity session (20 "indoor" activities).

details of each activity can be found later in this section):

1. Household cleaning – Vacuuming, sweeping, dusting.
2. Laundry – Folding clothes, ironing clothes.
3. Kitchen work – Peeling, slicing and dicing, grating, stirring, washing dishes, washing hands, washing vegetables.
4. Leisure activities – Watching TV.
5. Office-work – Talking on the phone, texting on the phone, using PC, writing with a pen.
6. Additional ambulatory activities – Walking upstairs, walking downstairs.
7. Other activities – Brushing teeth.

The data was all gathered indoors. Six monitors were available to gather the data with during the second session. The subjects included 17 males and 3 females, and ages ranging from 27 to 52. All subjects except for one were right-handed.

For both sessions, the data was gathered in a semi-controlled environment. The subjects were asked to perform a given activity and provided with equipment necessary to perform the activity. However, the subjects were not instructed on how exactly to perform the activity. Parameters such as how fast or what sequence of tasks they carried out while performing the activity was left for the subject to decide on. For some activities (e.g. watching TV) a time-limit was set.

Approval from the Human Research Ethics Advisory Panel (HREA) of the University of New South Wales Canberra was obtained for both data gathering sessions. The reference number of the HREA approval of the walking and running data gathering is *A-11-62* while that of the other 20 activities is *A-13-24*.

All the subjects were able-bodied volunteers. Signed consent was taken from each subject. For subjects who were minors at the time of the data gathering (the age of majority under Australian law at the time of data gathering is 18 years), signed consent of the subject's guardian was taken.

More details on the data gathering of each activity is given in the next subsections.

Walking

Subjects were asked to walk 20 meters on flat ground, turn and walk back. The subjects performed this activity three times, making a total distance of 120 meters recorded per subject-activity session. The subjects were asked to walk as fast as they felt was "normal", without specifying any time-limits or speed criteria.

Running

Similar to walking, subjects were asked to run 20 meters on flat ground, turn and run back. The subjects performed this activity three times, making a total distance of 120 meters recorded per subject-activity session. The subjects were asked for a "light run", without specifying any time-limits or speed criteria.

At the turn, some subjects took a large turn (hence did not slow down much), while some subjects took a sharp turn and hence slowed down for then turn then accelerated after the turn. These sections of the data were processed (included) with the data so as to include the inherent variations that exist when people take turns while running.

Brushing teeth

All the data was gathered in the same washroom, in front of a mirror and a wash-basin. The subject was provided a (new) toothbrush and a tube of toothpaste. Following which, the subject then applied some toothpaste onto the toothbrush. The subject then brushed his or her teeth. After brushing teeth, all subjects rinsed their mouths and the toothbrush (although not asked to do so). Data recording was stopped when the subject said that he/she was done brushing his/her teeth.

Dicing

All the data was gathered in the same kitchen area. The subject was provided a knife, a chopping board and a medium sized potato. This activity was recorded after the peeling activity in which the subject peels the potato, hence the potato was peeled. The subject was asked to dice the potato into pieces not more than 1cm by 1cm by 1cm in size. The subjects were not instructed on how to dice the potato. Different subjects used different methods to dice the potato. Most sliced it first, then sliced the slices into cuboid potato rods, then sliced the cuboid potato rods into small cubes. Two subjects cut a grid into the potato, then sliced out potato cubes from the grided potato. Data recording was stopped when the subject finished dicing.

Dusting

All the data was gathered in the same room. The subject was provided a feather duster and asked to dust an area of approximately five meters by seven meters. The dusting area contained six workstations with desktop computers, shelves with books and other lab equipment, a white board, a table with four chairs, a three-seater sofa and a filling cabinet. The subjects were asked to move around the room and were reminded to clean both high areas (e.g. shelves) and low areas (e.g. under the table). The subjects were not instructed on the order or the specific areas to clean. Data recording was stopped after 5 minutes of dusting.

Folding clothes

All the data was gathered in the same area. The subject was provided a pile of (clean) clothes. The clothes included shirts, T-shirts, trousers, dresses and several pairs of socks. The subject was then asked to fold the clothes. The subject was not instructed on how to fold the clothes. All subjects confirmed that they were familiar with how to fold the clothes. Data recording was stopped after 5 minutes of folding. In case, the subject folded all the clothes before the five minutes ran out, the researcher crumpled the clothes and asked the subject to repeat the process. Subjects were not required to finish the whole pile of clothes.

Grating

All the data was gathered in the same kitchen area. The subject was provided with a four-sided hand grater, a chopping board and a medium sized carrot. This activity was recorded after the peeling activity in which the subject peels the carrot, hence the carrot was peeled. The subject was asked to grate the carrot using any one of the grating widths available on the hand grater provided. The subjects were not instructed on how to grate the carrot. All the subjects placed the grater on the chopping board, holding the grater with their non-dominant hand while grating with their dominant hands. Data recording was stopped immediately after the subjects were finished grating the carrot.

Ironing clothes

All the data was gathered in the same area. The subject was provided a pile of (clean) clothes, an iron and an ironing board. The clothes included two shirts and two pairs of trousers. The subject was then asked to cold-iron the clothes. This is because due to the HREA specifications, the iron was not to be turned on during the data gathering. The subject was not instructed on how to iron the clothes. All subjects confirmed that they were familiar with how to iron clothes. Data recording was stopped either after 5 minutes of ironing or after ironing all the clothes. The subject was asked to iron one of the shirts first, then one of the trousers, then the other shirt and the other trouser. This was to ensure that were the 5 minutes to run out before the subject was finished ironing, at least data of ironing one shirt and one trouser was captured. All subjects managed to iron at least one shirt and one trouser within the 5 minutes.

Peeling vegetables

All the data was gathered in the same kitchen area. The subject was provided with a vegetable peeler, a medium sized carrot and a medium sized potato. The subject was asked to peel the potato first then the carrot. The subjects were not instructed on how to peel either the carrot or the potato. All the subjects peeled the carrot in seemingly the same way, however some subjects peeled the potato using long strokes from one end to another (like the carrot) while others made a continuous spiral peel (like while peeling an orange). Data recording was started immediately after the subjects started peeling the potato and stopped immediately after the subjects were finished peeling the carrot.

Stiring

All the data was gathered in the same kitchen area. The subject was provided with a wooden spoon and a saucepan. The subject was asked to pretend-stir the pan for 2 minutes. This is because due to the HREA specifications, the stove was not to be turned on during the data gathering. No further instructions were given to the subject. Data recording was stopped after 2 minutes of stiring.

Sweeping

All the data was gathered in the same room. The subject was provided a long handled broom and asked to sweep an area of approximately four meters by four meters. The area also contained a work desk and four chairs. The subject had to move the chairs around to sweep underneath them. The subjects were not instructed on how to sweep the area. Data recording was stopped after 2 minutes of sweeping.

Talking on the phone

The subject was asked to use the thigh monitor (mounted in the phone-like casing shown in figure 3.3) to make a pretend phone call. The subject was allowed to stand and walk around as he or she would if it were an actual call. In addition, the subject was instructed that sometime during the call, the person on the other side of the line would say a phone number to the subject, which the subject was to write down on a paper provided. Data recording was stopped immediately after the subject "hung up".

Texting on the phone

The subject was asked to type a pretend text message using his or her phone. In order to capture the sensors signals as observed from the phone, the thigh monitor (mounted in the phone-like casing shown in figure 3.3) was attached to the subject's phone. The subject was allowed to stand and walk around as he or she would if he or she found convenient. Data recording was stopped immediately after the subject was finished.

Using a PC

All the data was gathered at the same workstation and using the same PC. The subject was asked to find two advertised positions (employment)that fits him/her and email them to the author together with a brief (at least a paragraph) description of each role. This sequence of tasks was selected because it includes both the use of a mouse and keyboard, and involves both searching/browsing and synthesising typed information. Data recording was stopped immediately after the subject was finished.

Vacuuming

All the data was gathered in the same room. The subject was provided a vacuum and asked to vacuum an area of approximately four meters by four meters. The area also contained a sofa, a table and four chairs. The subject had to move the chairs around to vacuum underneath them. In addition, some random small items were dropped on the floor for the subject to pick up either before or during the vacuuming. The subjects were not instructed on how to vacuum the area. The subjects extended the power cable of the vacuum, plugged it into a wall socket and vacuumed the area. The subjects were asked to stop vacuuming after 5 minutes. Finally, the subjects unplugged the vacuum and place the vacuum back into it's original location. Data recording was stopped immediately after the subject was finished placing the vacuum back into it's original location.

Walking down stairs

Subjects were asked to walk down the stairs from the third floor of a building to the ground floor. The staircase went straight from one floor to a landing area, followed by an 180°turn on the landing area then more stairs to the next floor. The subjects were asked to walk as fast as they felt was "normal", without specifying any time-limits or speed criteria. Some subjects walked down slowly, some fast and some ran down the stairs two steps at a time. In addition, some subjects used the staircase handrail while some did not. Data recording was started immediately after taking the first step on the staircase, and stopped immediately after the last step of the staircase.

Walking up stairs

Subjects were asked to walk up the stairs from the ground floor of a building to the third floor. The staircase went straight from one floor to a landing area, followed by an 180°turn on the landing area then more stairs to the next floor. The subjects were asked to walk as fast as they felt was "normal", without specifying any time-limits or speed criteria. Some subjects walked up slowly, some fast and some ran up the stairs two steps at a time. In addition, some subjects used the staircase handrail while some did not. Data recording was started immediately after taking the first step on the staircase, and stopped immediately after the last step of the staircase.

Washing dishes

All the data was gathered at the same kitchen sink. The subject was provided a pair of cleaning rubber gloves (to prevent any water damage to the wrist monitors), a dish washing sponge and some liquid dish washing soap. In addition, the subject was also provided two plates, two cups, two spoons, two forks and two eating knives.

These were selected to represent the dishes resulting from a basic meal between two individuals. The subject was not instructed on how to wash the items or in which order to wash the items. All the subjects started by applying soap on the dishes and finished with rinsing the dishes and placing them on a dish rack next to the kitchen basin. Data recording was stopped immediately after the last item was placed on the kitchen drying rack.

Washing hands

All the data was gathered at the same kitchen sink. The subject was provided a bar of soap and asked to wash his or her hands. The subject was not instructed on how to wash his or her hands but was reminded to be thorough. Each subject turned on the tap, washed his/her hands and finally turned off the tap. Data recording was stopped immediately after turning off the tap.

Washing vegetables

All the data was gathered at the same kitchen sink. The subject was provided a medium-sized carrot and an unbrushed medium-sized potato. The subject was not instructed on how to wash the vegetables or in what order to wash them. Each subject turned on the tap, washed one vegetable, then washed the other vegetable and finally turned off the tap. Data recording was stopped immediately after turning off the tap.

Watching TV

All the data was gathered in the same location. The subject was asked to sit on a sofa and watch a TV mounted on the wall. A video was then selected with the help of the subject based on their preferences. This was performed to ensure that the subject was interested in the video watched on the TV. The videos were limited to videos of more than 5 minutes but less than 10 minutes. In addition, the subject was provided a remote to turn the TV on and off and some snacks to eat during the TV watching session. The subject turned the TV on, watched the video, and possibly snacked during the period. Data recording was stopped immediately after the video finished.

Writing

All the data was gathered at the same workstation. The subject was asked to hand write two paragraphs. The subject copied the first paragraph from another document, and was asked to write about his or her experiences on campus in the second paragraph. These tasks were selected because the first involves only writing while the second involves synthesis of information. It is possible that the two tasks produce different writing postures. No time-limits were set. Data recording was stopped immediately after the subject was finished.

3.4 Activity recognition data processing overview

Figure 3.7 gives an overview of the data processing performed to obtain activity classification results. Within different sections of the thesis, slight variations of this process are used to obtain classification results. The classification results are then used to analyse the impact of the varied factors upon success-rates.

Sampling of data was performed at 128Hz to obtain accelerations, rotational velocities and orientation values from six monitors mounted on different locations on the subject's body. More details about the data gathering process are given in section 3.3. Once the data has been sampled, an optional task of preprocessing data at a record level (i.e. preprocessing performed to all data gathered in one subject's activity session) might be performed depending on the analysis being performed. An example of record-level preprocessing is the application of random rotations to

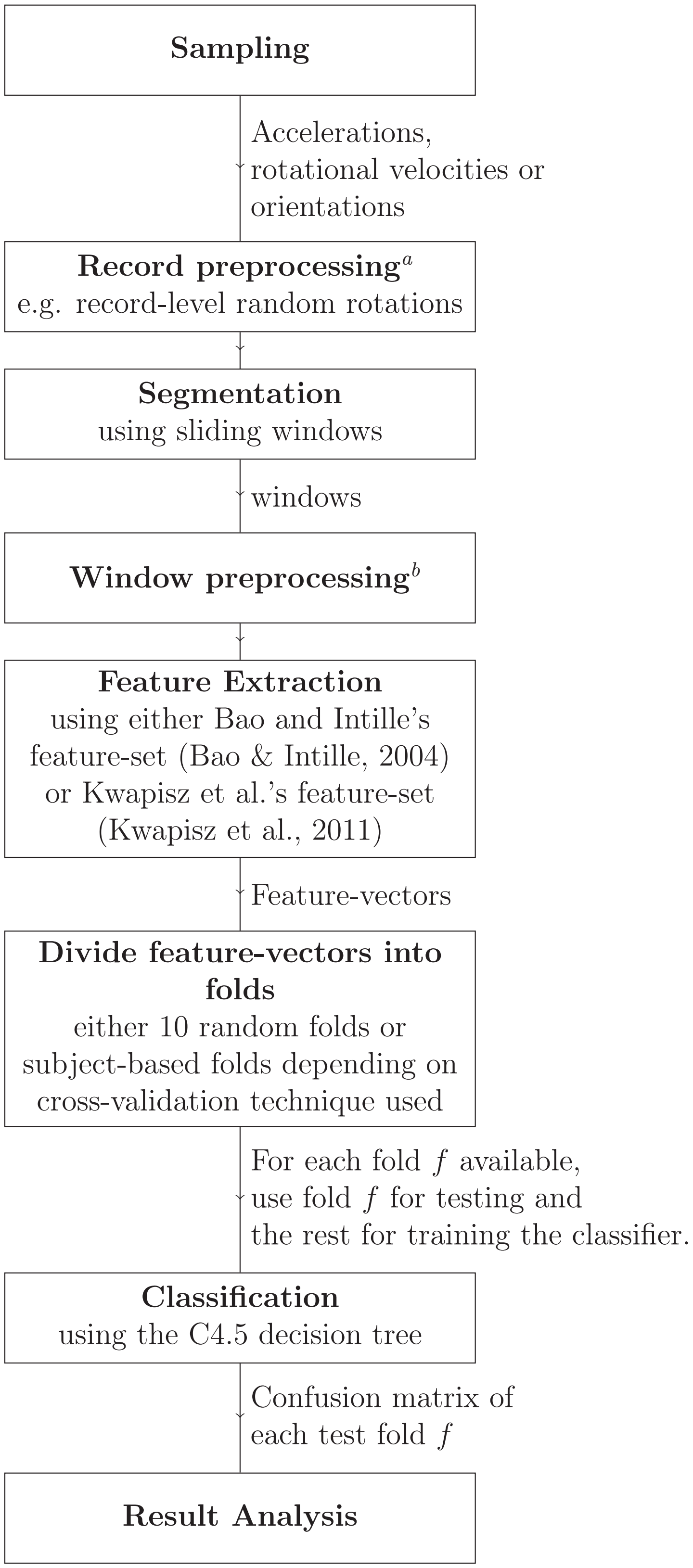


Figure 3.7: An overview of the activity recognition data processing performed.

[a] Optional step

[b] Includes downsampling

the whole record. This is followed by segmentation, where sliding windows are used to divide the record into smaller sequences (windows). The parameters of the sliding window differ depending on the analysis being performed. The windows can then be further preprocessed. This includes downsampling the window. Next, features are extracted using either Bao and Intille's feature-set (Bao & Intille, 2004) or Kwapisz et al.'s feature-set (Kwapisz et al., 2011) (refer to section 3.5 for more details).

All the feature-vectors generated are then divided into folds. Either 10-fold or remove-one-subject cross-validation is used depending on the analysis performed (refer to section 3.6 for more details). If 10-fold cross-validation is used, the feature-vectors are divided into 10 random folds. If remove-one-subject cross-validation is used, the feature-vectors are divided into folds based on the subject the data was gathered from. Next, the folds are used to train and test a C4.5 decision tree classifier (Quinlan, 1993). For each of the folds obtained, the classifier is tested with that fold after being trained with the other folds. This results in one confusion matrix for each fold. The confusion matrixes are then used for analysis to answer the relevant research questions.

It should be noted that the definition of feature extraction and feature-set within the context of this thesis might be different from the norm. Refer to the glossary for the definitions.

3.5 Feature-sets used

The research questions possed in the thesis require feature-extraction and classification. For feature-extraction, two different feature-sets were selected and used to answer the research questions so as to give more weight on the conclusions made.

The feature-sets used are based on those proposed by Bao (2003) and Kwapisz et al. (2011). The feature-sets from these publications were selected because:

1. The feature-sets do not require input from non-inertial sensors. For example,

the algorithm proposed by Saponas et al. (2008) requires a footpod, Ermes and Juha (2008) requires a respiratory effort sensor, and Quwaider and Biswas (2008) requires the signal RSSI from the communications with the wearable sensors.

2. The feature-sets proposed in the publications are explained well enough to be implementable.
3. The algorithms precisely describe the terms used. This contrasts with some proposed feature-sets which require the implementer to search for the optimum parameters in order to achieve the published success rate of the algorithm.
4. The feature-sets performed well in our initial testing.

Bao and Intille proposed their algorithm for activity recognition systems that make use of a body-sensor network such that data is processed from accelerometers mounted on different locations on the subject. The feature-set derived from Bao and Intille's algorithm is given in table 3.3.

All of the features from Bao and Intille's feature-set, except for the Pearson product-moment correlation coefficient, can be computed from the frequency components of the signal. Hence the feature-set can be described as primarily based on frequency-domain features.

Kwapisz et al. proposed their algorithm for activity recognition systems that make use of the smart phone as a sensing node such that the only data comes from the phone's accelerometer, which might be carried in one of many phone carry locations on the subject's body and in any orientation. The feature-set derived from Kwapisz et al.'s algorithm is given in table 3.4.

All the features of Kwapsiz et al.'s feature-set are computed in the time-domain. Although Kwapisz et al.'s feature-set is designed for use in smart phone-based activity recognition, with the exception of the average magnitude of the sample vectors, the values of other features change with the orientation of the phone because they are specific to the accelerometer's axis to which they are measured.

Table 3.3: Feature-set derived from Bao (2003). x,y and z represent x-axis, y-axis and z-axis signal in the time-domain while X,Y and Z represent the same in the frequency-domain.

Feature	**Definition**
Mean of each axis of the signal	X_1 Y_1 Z_1
Information entropy of each axis of the signal (excluding first bin)	$-\sum_{i=2}^{N}(p_i \log_2 p_i), \quad p_i = \frac{\lvert X_i\rvert^2}{\sum_{j=2}^{N}\lvert X_j\rvert^2}$ $-\sum_{i=2}^{N}(p_i \log_2 p_i), \quad p_i = \frac{\lvert Y_i\rvert^2}{\sum_{j=2}^{N}\lvert Y_j\rvert^2}$ $-\sum_{i=2}^{N}(p_i \log_2 p_i), \quad p_i = \frac{\lvert Z_i\rvert^2}{\sum_{j=2}^{N}\lvert Z_j\rvert^2}$
Signal energy of each axis of the signal (excluding first bin)	$\frac{1}{N-1}\sum_{i=2}^{N}(\lvert X_i\rvert^2)$ $\frac{1}{N-1}\sum_{i=2}^{N}(\lvert Y_i\rvert^2)$ $\frac{1}{N-1}\sum_{i=2}^{N}(\lvert Z_i\rvert^2)$
Pearson product-moment correlation coefficient of each pair of sensor axes	$\frac{1}{N}\sum_{i=1}^{N}\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}$ $\frac{1}{N}\sum_{i=1}^{N}\frac{(x-\bar{x})(z-\bar{z})}{\sigma_x\sigma_z}$ $\frac{1}{N}\sum_{i=1}^{N}\frac{(y-\bar{y})(z-\bar{z})}{\sigma_y\sigma_z}$

3.6 Cross-validation techniques

A review and comparison of various cross-validation techniques performed in the literature can be found in section 2.6.

If we are to compare the impact of various parameters in the activity recognition system, we wish to evaluate the best-case-scenario. Error rates due to inter-subject variability can impact the analysis leading to either unclear results or results that don't reflect the potential of the parameter. For this reason, many sections in the chapter use 10-fold cross-validation (which reflects results achievable by a hybrid model) instead of remove-one-subject cross-validation (which reflects results achievable by an impersonal model).

Both cross-validation techniques were used in the analysis, but with the exception

Table 3.4: Feature-set derived from Kwapisz et al. (2011). x,y and z represent x-axis, y-axis and z-axis signal in the time-domain.

Feature	Definition
Mean of each axis of the signal	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ $\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$
S.D. of each axis of the signal	$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ $\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$ $\sigma_z = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})^2}$
Average absolute difference from mean of each axis	$\frac{1}{N}\sum_{i=1}^{N} \lvert x_i - \bar{x}\rvert$ $\frac{1}{N}\sum_{i=1}^{N} \lvert y_i - \bar{y}\rvert$ $\frac{1}{N}\sum_{i=1}^{N} \lvert z_i - \bar{z}\rvert$
Average time in milliseconds between peaks of each axis	$\frac{1}{M}\sum_{i=1}^{M} p_i, \quad p = \{i \mid (x_i > x_{i-1}), (x_i > x_{i+1})\},\ M = length(p)$ $\frac{1}{M}\sum_{i=1}^{M} p_i, \quad p = \{i \mid (y_i > y_{i-1}), (y_i > y_{i+1})\},\ M = length(p)$ $\frac{1}{M}\sum_{i=1}^{M} p_i, \quad p = \{i \mid (z_i > z_{i-1}), (z_i > z_{i+1})\},\ M = length(p)$
Average of the magnitudes of sample vectors	$\frac{1}{N}\sum_{i=1}^{N} \sqrt{x_i^2 + y_i^2 + z_i^2}$
The distribution of the signal values into 10 bins	The range of values of each axis is divided into 10 equal sized bins. The fraction of samples within the window that falls within each bin is then used as a feature

of section 4.4 which deals with the impact of window overlaps, results of other sections using 10-fold were found to have less variability, higher statistical significance and therefore clearer and easier to derive conclusions from than those of remove-one-subject cross-validation. Further analysis on the comparison of success-rates obtained using 10-fold cross-validation and those obtained from remove-one-subject cross-validation is found in section 6.4.

3.7 Methodology for determining impact on success-rates

In order to compare the impact of various data processing techniques on the activity classification success-rates obtained, the same data is processed twice: once without the data processing technique in question, and once with the data processing technique. The resulting success-rates obtained from each data set are then subtracted from each other. A (Gaussian) model is then fitted onto the differences obtained and the fitted model is used to ascertain whether or not the technique in question impacts success-rates.

This methodology was selected because:

1. Due to cross-validation, a slightly different activity classification success-rate is obtained from each test fold. The success-rates were observed to fit a Gaussian model. However, the success-rates obtained are only a small sample. Hence, to obtain more accurate results, a Gaussian model is fitted using Maximum Likelihood Estimation. The estimated population mean and standard deviation are then extracted from the fitted model.
2. Using the fitted model, additional information can be extracted. In particular, we are interested in the probability that application of the data processing technique results in a higher success-rate in a sample than when the data processing technique is not applied. Figure 3.8 illustrates why it is important to compute the mentioned probability.

Figure 3.8 shows a hypothetical example of the differences in success-rates obtained from process a, b and c applied to the same samples. The mean difference in the success-rates obtained by process a and b is 0.5 and standard deviation 0.5. The mean difference in the success-rates obtained by process a and c is 1.0 and standard deviation 2.0.

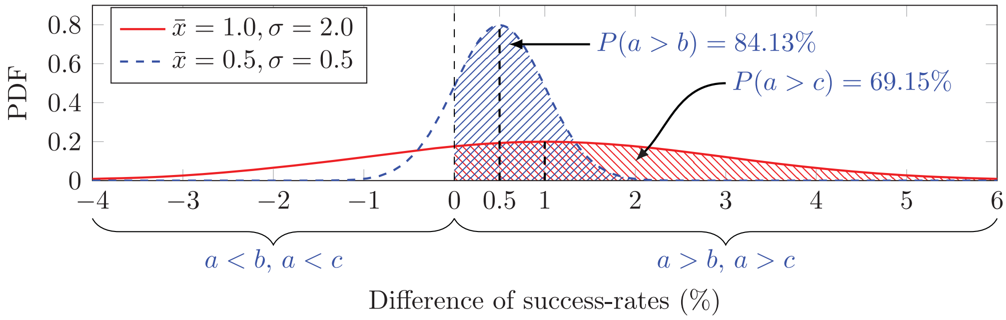


Figure 3.8: A hypothetical example of differences in success-rates obtained from process a, b and c applied to the same samples. One set of results (blue) is the difference between a and b (mean=0.5, standard deviation=0.5) and the other set (red) is the difference between a and c (mean=1.0, standard deviation=2.0). The percentage of the area above $a - b = 0$ is higher than that above $a - c = 0$.

When comparing between a, b and c, it can be observed that a results in higher success-rates than both b and c. In addition, the mean difference between a and c is higher than the difference between a and b. However, the differences between a and c also have a higher variance than the differences between a and b.

From the mean differences we can observe which process results in higher success-rates. However, we can not observe how consistently the process is in resulting in higher success-rates. In the example in figure 3.8, it can be observed that process a results in higher success-rates than process b more consistently than process a results in higher success-rates than process c. This can be observed from the variance of the distribution of the differences in success-rates.

By combining the mean and the variance of each distribution of the differences in success-rate, it is possible to assign a value to how consistent one process results in higher success-rates than the other process: by computing the likelihood of obtaining higher results using the former process than the latter process. The likelihood has an additional advantage in that it can be compared across processes. It should be noted that the likelihood is still highly linked to the mean: a likelihood $> 50\%$ of process a resulting in higher success-rates than process b means that the mean success-rates of a are higher than the mean success-rates of b.

4 Analysis of the impact of data capture and data processing parameters upon activity classification accuracy

4.1 Introduction

In this chapter, fundamental settings that impact activity classification accuracy are explored. The settings explored are: minimum sampling frequencies, window length and window overlap. The chapter is organised as follows:

First, the lowest sampling frequencies, sampling above which result in no further increase in activity classification success-rates, are explored. Using these frequencies, the sampling requirements of accelerations, rotational velocities and orientations are analysed.

Next, the relationship between the length of windows used to extract feature-vectors and the resultant activity classification success-rates is explored. The relationship is modelled and the model is extended to find the longest window length, above which, no further increase impacts activity classification success-rates. In the discussion section, an explanation of the observed relationship between window lengths and resultant success-rates is offered.

Finally, the impact of window overlaps on the resultant activity classification success-rates is ascertained for two cross-validation methods. Where an impact is found, the relationship between window overlaps on both mean and variance of the resultant activity classification success-rates is modelled. In the discussion section, an explanation of the observed relationship between window overlaps and the means and variances of the resultant activity classification success-rates is offered.

4.2 Impact of sampling frequency on activity recognition accuracy

The lower the sampling frequency used in an activity recognition system, the lower the amount of data that has to be processed for feature extraction. This reduces the computational resources (like memory, computation time and battery power) required for activity recognition. In wearable activity recognition systems and smart phones, such resources are likely to remain limited for the foreseeable future.

In this section, the impact of the sampling frequency on the classification success-rate of the gathered activities is studied.

Oversampling results in no improvement in classification success-rates. Hence, it is important to define the frequency at which a further increase in sampling frequency results in no further improvement in success-rates.

To that end, we define the Minimum Efficient Sampling (MES) frequency. MES frequency is the lowest sampling frequency to achieve an activity classification success-rate that has a 95% chance of being independent of the sampling frequency from the data of a sensor mounted on a particular location on the body.

In other words, given a location on the body, data sampled at the MES frequency on the location, results in success-rates that have a 95% chance of being in the same distribution as success-rates obtained from the same location using frequencies higher than the MES frequency.

We then use the MES frequencies to answer the following research questions:

1. **Which *source* derived features have the lowest MES frequency on**

average?

2. **In addition, which *source* derived features have the highest MES frequency on average?**

By analysing the differences in MES frequencies extracted from acceleration, rotation velocity and orientation derived features of all the activities and the monitors used, we can learn which *source* derived features result in the lowest and highest MES average frequency.

The order of the *sources* in terms of increasing MES average frequency can indicate the order of the sampling frequency requirements of the *source* in terms of increasing sampling frequency. Knowing this could aid an activity recognition system developer in selecting which *source* to utilise depending on the constraints placed on the system (like power consumption and system accuracy).

The next section presents the methodology used to compute the MES frequencies and to answer the research questions.

Methodology

The original data was downsampled to frequencies ranging from 1Hz to 128Hz in steps of 1Hz. The frequency downsampled-to is referred to as the downsampling frequency henceforth. Appropriate low pass filtering was performed prior to each downsampling. Algorithm 1 further elaborates on this process.

A sliding window of 10 seconds with 50% overlap was used.

Window lengths of 10 seconds were used because 10 seconds was observed to be the maximum window length used in activity recognition literature, having only been used by Kwapisz et al. (2011) and Patel et al. (2009). In the analysis of activity recognition literature performed by Lockhart and Weiss (Lockhart & Weiss, 2014), window lengths reported to have been used in activity recognition literature were observed to have a median of 3 seconds and the maximum window length they

Algorithm 1 Process the time-series D resulting into a set of time-series, that contains one time-series for each downsampling frequency.

procedure DownsampleTimeSeries(D)
    $Results \leftarrow []$ ▷ List to hold resulting downsampled time-series
    $n \leftarrow |D|$ ▷ Get number of samples in D
    $freq \leftarrow CompBinFreqs(128, n)$ ▷ Compute FFT bin frequencies for n bins and to the original sampling frequency of 128Hz
    $fft \leftarrow FFT(D)$ ▷ Convert D to frequency domain

    for all $f \in [128, 127, ..., 1]$ **do**
        for all $i \in [0, n)$ **do** ▷ Clear all frequencies equal or above
            if $freq_i \geq \frac{f}{2}$ **then** ▷ $\frac{f}{2}$ in the FFT results
                $fft_i \leftarrow 0$
            end if
        end for

        $filtered \leftarrow IFFT(fft)$ ▷ Convert filtered data back to time domain

        $m \leftarrow \left\lfloor \frac{fn}{128} \right\rfloor$ ▷ Compute resulting number of samples

        ▷ Average contiguous groups of $\frac{128}{f}$ samples.
        for all $i \in [0, m)$ **do**
            $s \leftarrow \left\lfloor \frac{in}{m} \right\rfloor$ ▷ First sample of D to average
            $e \leftarrow \left\lfloor \frac{(i+1)n}{m} \right\rfloor$ ▷ Last sample of D to average
            $r_i \leftarrow Mean(filtered_{[s...e)})$ ▷ Compute downsampled sample.
        end for

        $Results$.add(r) ▷ Add to result list
    end for

    return $Results$
end procedure

observed was 10 seconds.

A 50% window overlap was used because it was observed that 50% window overlaps are common within the literature review having been used by Bao and Intille (2004); Figo et al. (2010); He et al. (2008); Krishnan and Panchanathan (2008); Kunze et al. (2005); Preece, Goulermas, Kenney, and Howard (2009); Ravi et al. (2005); Shoaib et al. (2014) and Sun et al. (2010). However, other window overlaps also exist in the literature review including: no overlap ((Kwapisz et al., 2011)), 20% ((Reiss, 2014)), 25% overlap ((Henpraserttae et al., 2011)), 33% overlap ((Lester et al., 2005)).

Features were then extracted as explained in the respective feature-set's paper (Bao and Intille (2004) and Kwapisz et al. (2011), refer to section 3.5 for more details). In addition, a Hamming window was applied to each window before extracting the frequency-domain features in Bao and Intille's feature set.

Finally, classification was performed using the J48 decision tree from the WEKA toolkit. The J48 decision tree is an implementation of the C4.5 algorithm (Hall et al., 2009). The C4.5 decision tree was found to perform best by Bao and Intille (Bao & Intille, 2004), and second best by Kwapisz et al. (Kwapisz et al., 2011).

Experiments reported in this section were performed using 10-fold cross-validation.

The resulting success rates as a function of the downsampling frequency were observed to be noisy. Ideally, success-rates as a function of downsampling frequencies are expected to monotonically increase with increase in downsampling frequency. However, while this was the overall trend observed in the obtained success-rates, local seemingly-random fluctuations were also observed (see figure 4.1 for an example).

To extract the MES frequency from the obtained success-rates as a function of downsampling frequency, a two step process is followed:

1. An equation that represents the expected model of results was fitted on to the observed results.

The expected model is based on the assumption that the success-rate obtained for classifying an activity would be lowest when sampling at 0Hz (i.e. sampling the mean of the signal only), increasing as the sampling frequency increases until a sampling frequency which is higher than twice the highest frequency component that can identify the activity (at frequency F), then stop changing and maintain that success-rate independent of any further increases of the sampling frequency.

Fitting an equation that represents this model and using the estimated values of the model instead of the observed values reduces the impact of the local success-rate fluctuations on the analysis. The equation that summarises this behaviour expected of success-rates y as a function of downsampling frequencies x is given as equation 4.1. See figure 4.1 for an example of equation 4.1 fitted onto data.

$$y' = \begin{cases} \dfrac{B-C}{F}x + C, & \text{if } x < F \\ C, & \text{otherwise} \end{cases} \tag{4.1}$$

where x = downsampling frequency in the range [1,128],

y' = estimated success-rate of data downsampled to frequency x,

C = estimated success-rate of data dowsampled to 0Hz,

F = lowest frequency at which the success-rate of data is unaffected by the downsampling frequency.

$$B = \frac{1}{(128 - F + 1)} \sum_{i=F}^{128} y_i$$

The equation divides the [1Hz,128Hz] frequency range into two: a lower range that is estimated to impact success-rates, and a higher range that is estimated not to impact success-rates. The frequency that separates these two ranges is denoted as F. The lower range that is estimated to impact success-rates is modelled as a linear function of positive slope. The upper range that is estimated not to impact success-rates is modelled as a constant success-rate.

2. The MES frequency was extracted as being less than the F frequency, and whose estimated success-rate (according to the fitted equation) is within two standard deviations of the mean of the success-rates of the frequency range above F (see figure 4.1 for an example).

   Extracting the MES frequency based on the likelihood of its estimated success-rate originating from the distribution of success-rates from downsampling frequencies above F, allows for the selection of an MES frequency such that the characteristics of success-rates from frequencies below it are distinct from those above it.

   This is especially important for cases where the slope of success-rates below frequency F is small (see figure 4.2 for an example). In such cases, even though equation 4.1 is fitted correctly by definition, the characteristics of success-rates below F and above F are largely similar.

Figure 4.1 elaborates on the fitting of equation 4.1 to some examplar data. The example shows a general increase in success-rate with increase in downsampling frequency to approximately 33Hz. Thereafter, the success-rates remain generally constant with increase in frequency. The frequency F was computed as 33Hz and MES frequency as 29Hz.

To fit the equation, values of F in the range of [1Hz,128Hz] in steps of 1Hz are tested to find the value of F that results in the fitted equation with the highest R^2. A non-linear least squares method was used to fit the equation.

Cases exist where the slope of success-rate to downsampling frequency is so low that the range of frequencies immediately below F also forms part of the distribution of success-rates that are estimated not to effect the sampling frequency. The extreme

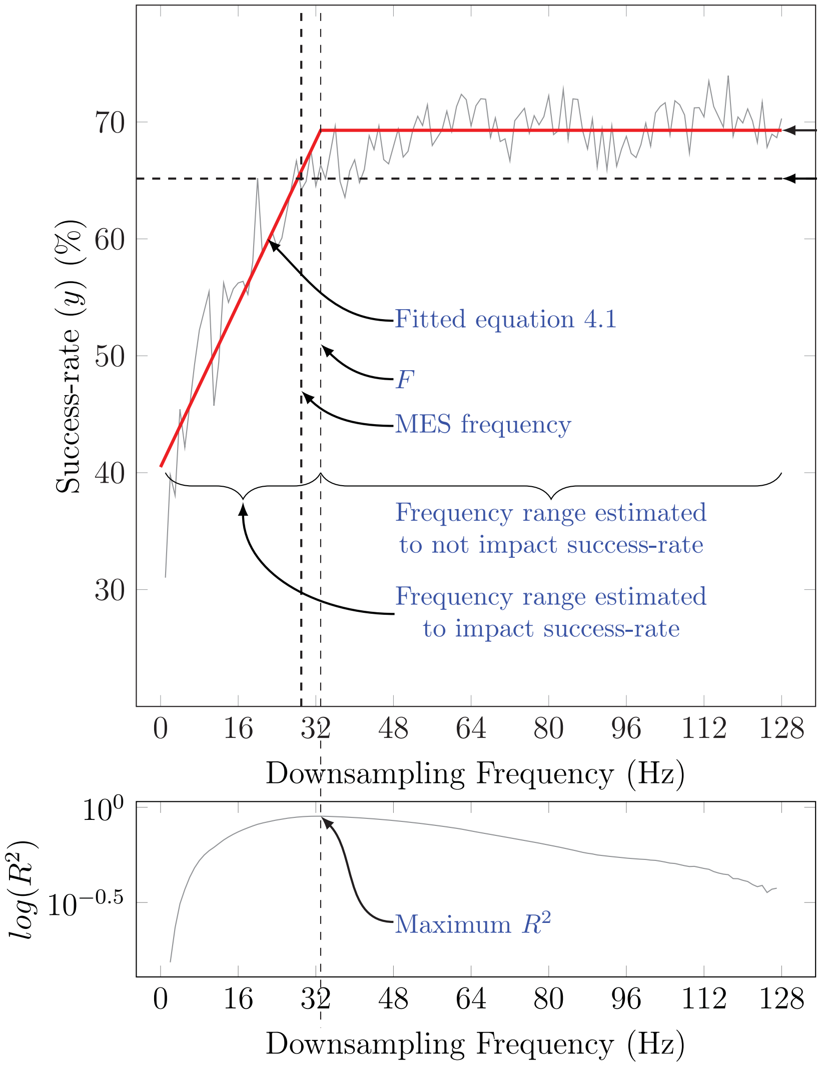


Figure 4.1: Success-rate as a function of downsampling frequency exemplifying results obtained from classification of data downsampled to frequencies in the range [1Hz,128Hz] in steps of 1Hz. This particular example shows a general increase in success-rate with increase in downsampling frequency to approximately 33Hz. Thereafter, the success-rates remain generally constant with increase in frequency. The chart also presents the fitted equation 4.1, the computed F frequency of 33Hz and MES frequency of 29Hz. The lower graph is a semilog plot of R^2 values obtained for each frequency value while fitting equation 4.1.

case of this is when an activity is recognised with the highest success-rate $\leq$ 1Hz. In that case, the success-rates would not experience the linear growth in frequencies below F assumed in the model. Figure 4.2 gives an example of one such case where the highest success-rate is at 1Hz and there is no apparent general increase in success-rate with increase in downsampling frequency.

Hence, we define the MES frequency to be the lowest frequency with an estimated success-rate that has a 95% chance of belonging to the same distribution as $y_{[F..128]}$ (the success-rates found to be independent of the downsampling frequency). This would include any frequencies immediately below F that have a 95% chance of resulting in a success-rate that belongs to the distribution of $y_{[F..128]}$.

After fitting equation 4.1 to the data we can use the mean ($\overline{y_{[F..128]}}$) and standard deviation ($\sigma_{[F..128]}$) of $y_{[F..128]}$ to find the lowest estimated success-rate that is $\geq$ $\overline{y_{[F..128]}} - 2\sigma_{[F..128]}$

The process of extracting the MES frequency from activity classification success-rates is elaborated on in algorithm 2.

The process was repeated for accelerations, rotational velocities and orientations.

To compare the MES frequencies extracted from accelerations, rotational velocities and orientations, the MES frequency of each *source* is extracted for each activity and monitor.

The MES frequencies of any two *sources* can then be compared in a method based on a paired two-sample T-Test. A paired two-sample T-Test tests the null hypothesis that the two collections of samples come from normal distributions of equal means but unknown variances. It does this by subtracting samples of one collection from the other and checking whether the resultant distribution has a mean of zero at the given significance level. However, in this case, we are interested in which *source* has greater MES frequencies than the other.

Therefore, after running a paired two-sample T-Test (α = 0.05) on the MES frequencies of two *sources* to check that the MES frequencies of the two *sources* are

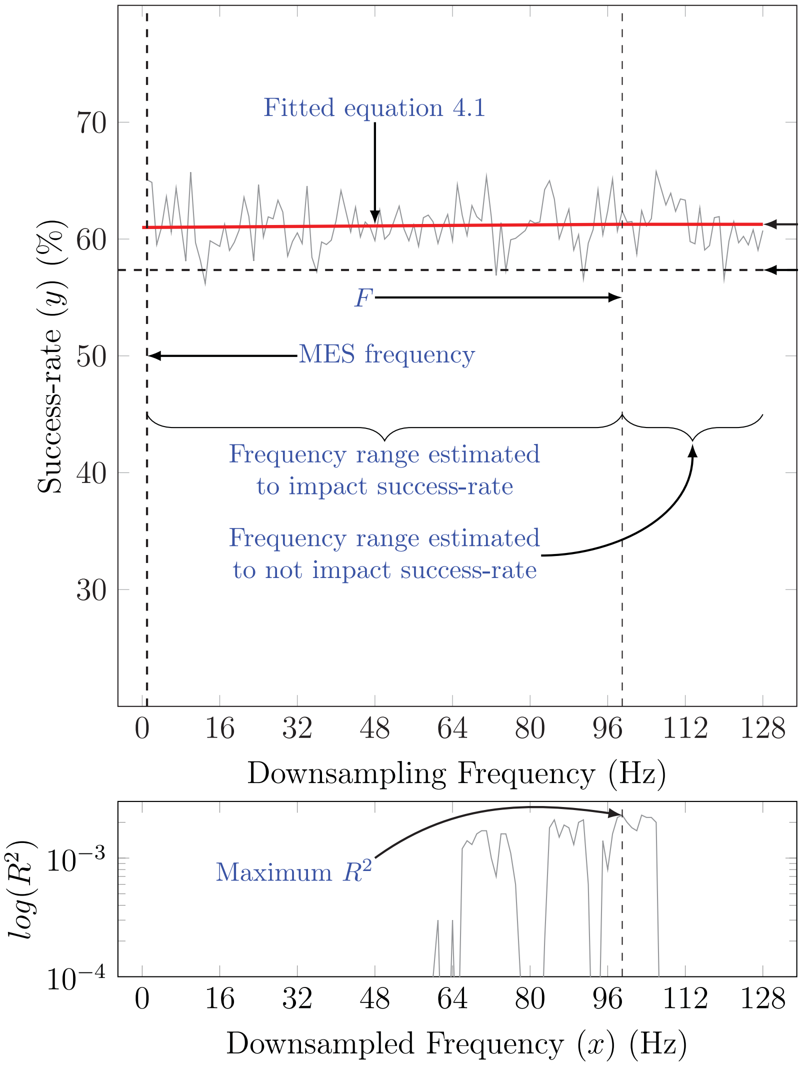


Figure 4.2: Mean success rate as a function of downsampling frequency exemplifying results obtained from classification of data downsampled to frequencies in the range [1Hz,128Hz] in steps of 1Hz. This particular example shows no apparent general increase in success-rate with increase in downsampling frequency. The chart also presents the fitted equation 4.1, the computed F frequency of 99Hz and MES frequency of 1Hz. The lower graph is a semilog plot of R^2 values obtained for each frequency value while fitting equation 4.1.

Algorithm 2 Compute the MES frequency given the success-rates obtained from classifying data downsampled to frequencies in the range [1Hz,128Hz] in steps of 1Hz.

procedure COMPUTEMESF($test\_results$)
▷ $test\_results$ are classification results of a particular activity, monitor and *source* and include results of classifying data downsampled to frequencies in the range [1Hz,128Hz] in steps of 1Hz for either Bao and Intille's feature-set or Kwapisz et al's feature-set.
    $bsf\_R^2 \leftarrow 0$ ▷ Best so far R^2 from curve fitting.
    $F \leftarrow 0$
    $A \leftarrow 0$
    $M \leftarrow 0$
    for all $f \leftarrow$ 1Hz to 128Hz in steps of 1Hz **do**
        $c \leftarrow Mean(y_{[f..128]})$ ▷ Compute mean of success-rates obtained from frequencies f to 128Hz
        Fit equation 4.1 to data.
        if $R^2 > bsf\_R^2$ **then** ▷ Find f that results to the highest R^2 value.
            $bsf\_R^2 \leftarrow R^2$
            $F \leftarrow f$
            $A \leftarrow a$
            $M \leftarrow m$
        end if
    end for

    $\sigma_{[F..128]} \leftarrow StandardDeviation(y_{[F..128]})$
    $\overline{y_{[F..128]}} \leftarrow Mean(y_{[F..128]})$ ▷ Obtain the mean and standard deviation success-rates that are independent of the data downsampling frequency.
    for all $i \leftarrow$ 1Hz to F in steps of 1Hz **do**
        if $y'_i \geq \overline{y_{[F..128]}} - 2 * \sigma_{[F..128]}$ **then**
▷ Check if the estimated success-rate is within two standard deviations of the mean of success-rates that are independent of the data downsampling frequency.
            return i ▷ Return frequency i if y'_i does not belong to the distribution.
        end if
    end for
    return 0Hz ▷ Return 0Hz if no MES frequency was found in the range [1Hz,128Hz].
end procedure

not equal, the mean signed difference between the two sources is used to find out which *source* is larger than the other.

A Gaussian model is then fitted onto the histogram of the difference of the two *sources* so as to extract a mean and standard deviation that is more representative of the population mean and standard deviation of the difference of the two sources.

Algorithm 3 was used to compute the differences in MES frequencies of one *source* to another *source*.

Algorithm 3 Compute the difference in MES frequencies of one *source* to another *source* for all activities and monitors given.

procedure MesfDiff(C A, M, $S1$, $S2$)
▷ C are classification results and include results of classifying data downsampled to frequencies in the range [1Hz,128Hz] in steps of 1Hz for either Bao and Intille's feature-set or Kwapisz et al's feature-set.
▷ A is a list of activities
▷ M is a list of monitors
▷ $S1$ is the first *source*
▷ $S2$ is the second *source*

$Results \leftarrow []$ ▷ List to hold resulting values

for all $test\_results \in C$ **do**
for all $a \in A$ **do**
for all $m \in M$ **do**
$d1 \leftarrow test\_results_{activity=a,monitor=m,source=s1}$
▷ Extract results of activity a, monitor m and source $s1$ from the test results
$d2 \leftarrow test\_results_{activity=a,monitor=m,source=s2}$
▷ Extract results of activity a, monitor m and source $s2$ from the test results
$f1 \leftarrow ComputeMesF(d1)$
$f2 \leftarrow ComputeMesF(d2)$
$Results$.add$(f1 - f2)$
end for
end for
end for
Perform One-Sample T-Test to check whether *Results* distribution is centered at zero.
Fit a Gaussian model onto *Results* using Nonlinear Least Squares method.
Extract mean and standard deviation from fitted model.
end procedure

The next section presents results obtained from this analysis.

Results

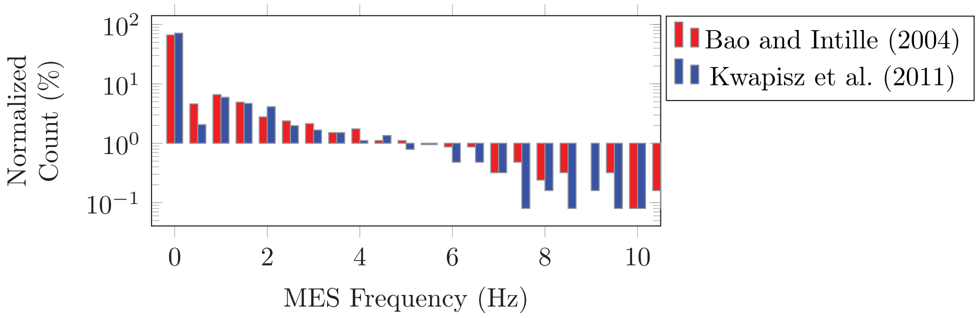
The complete set of plots showing success-rates in classifying each activity plotted against the frequency the data was downsampled to are presented in appendix A. In addition, the fitted equation 4.1 is also shown on each of the figures. Results from accelerations, rotational velocities and orientations, from each of the monitors available, and from the two feature-sets used (Bao and Intille's and Kwapisz et al.) are given.

The extracted MES frequencies obtained using data derived from activities, monitors, *sources* and from the two feature-sets used (Bao and Intille's and Kwapisz et al.), can be found in appendix B.

Histograms of the MES frequencies obtained using data derived from all activities, all monitors and from the two feature-sets used (Bao and Intille's and Kwapisz et al.) are shown in figure 4.3, figure 4.4 and figure 4.5 for accelerations, rotational velocities and orientations respectively.

The histograms show that a large percentage of the data for all three *sources* is concentrated at 0Hz. This is because many combinations of body-locations and activities result in MES frequencies of 0Hz. This indicates that higher frequencies have no impact on the classification success-rates on those combinations of body-locations and activities.

However, higher MES frequencies are also seen for each of the *sources*. Of the three *sources*, it can be observed that the area under the graph above 0Hz is larger for the rotational velocities derived histogram than for either the accelerations or orientations derived histograms. This indicates that generally rotational velocities resulted in higher MES frequencies than either accelerations or orientations. In addition, it can also be observed that the area under the graph above 0Hz is smaller



	Bao	***Kwapisz***
Median	0.01	0.01
Inter-Quartile Range	1.08	0.90
$\bar{x}$	0.91	0.81
σ	0.80	2.30

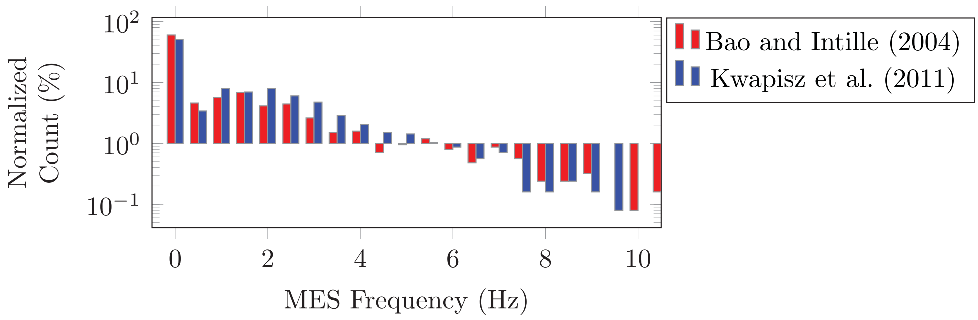
Figure 4.3: Distribution of MES frequencies extracted from classification results of accelerations for each activity and monitor. Results from both Bao and Intille's feature-set as well as Kwapisz et al.'s feature-set are shown.

for the orientations derived histogram than for either the accelerations or rotational velocities derived histograms. This indicates that generally orientations resulted in lower MES frequencies than either accelerations or rotational velocities. A better analysis will be conducted next where for each activity and body-location, the MES frequency obtained for each *source* is compared against each other *source*. This results in more accurate conclusions of how different the MES frequencies of the three *sources* are.

In the following sections, the results of comparing MES frequencies extracted from each *source* to each of the other *sources* are presented and then a conclusion is made about the following research question.

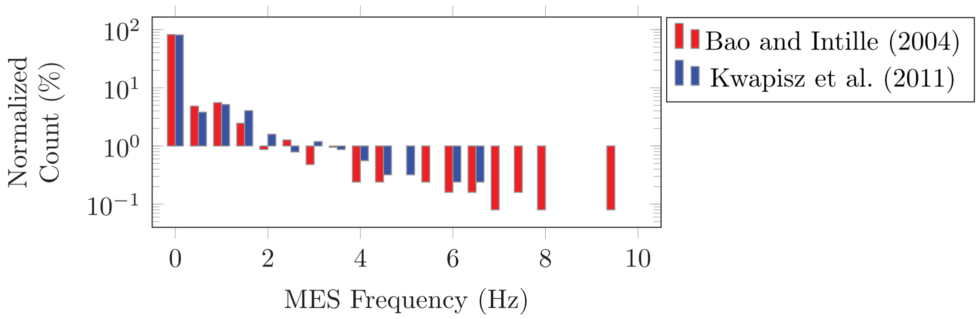
Do acceleration features require higher MES frequencies than orientation features?

The distribution of the results of applying algorithm 3 on accelerations and orientations using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown in figure 4.6.



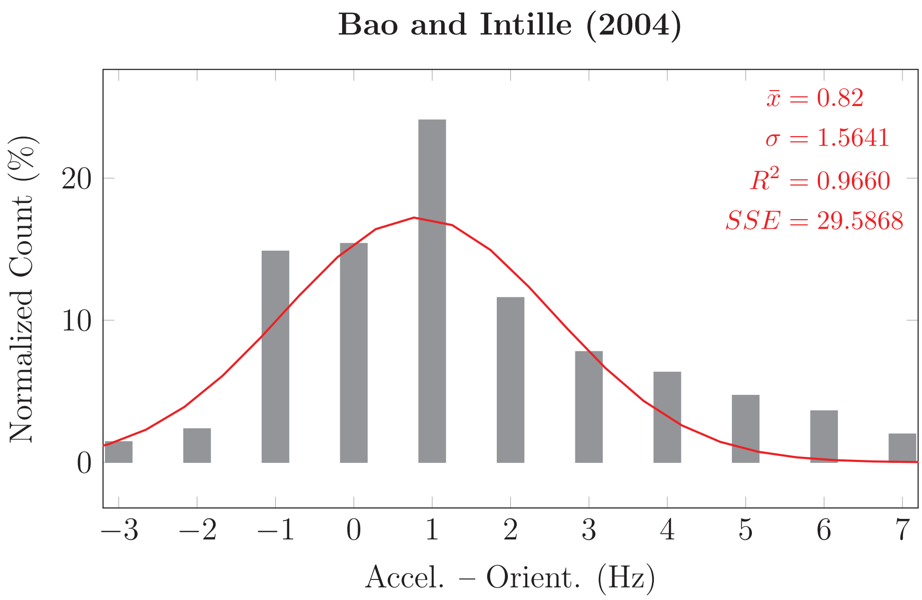
	Bao	***Kwapisz***
Median	0.00	0.17
Inter-Quartile Range	1.56	2.10
$\bar{x}$	1.36	1.28
σ	3.34	1.83

Figure 4.4: Distribution of MES frequencies extracted from classification results of rotational velocities for each activity and monitor. Results from both Bao and Intille's feature-set as well as Kwapisz et al.'s feature-set are shown.



	Bao	***Kwapisz***
Median	0.00	0.00
Inter-Quartile Range	0.00	0.00
$\bar{x}$	0.30	0.33
σ	0.92	0.89

Figure 4.5: Distribution of MES frequencies extracted from classification results of orientations for each activity and monitor. Results from both Bao and Intille's feature-set as well as Kwapisz et al.'s feature-set are shown.



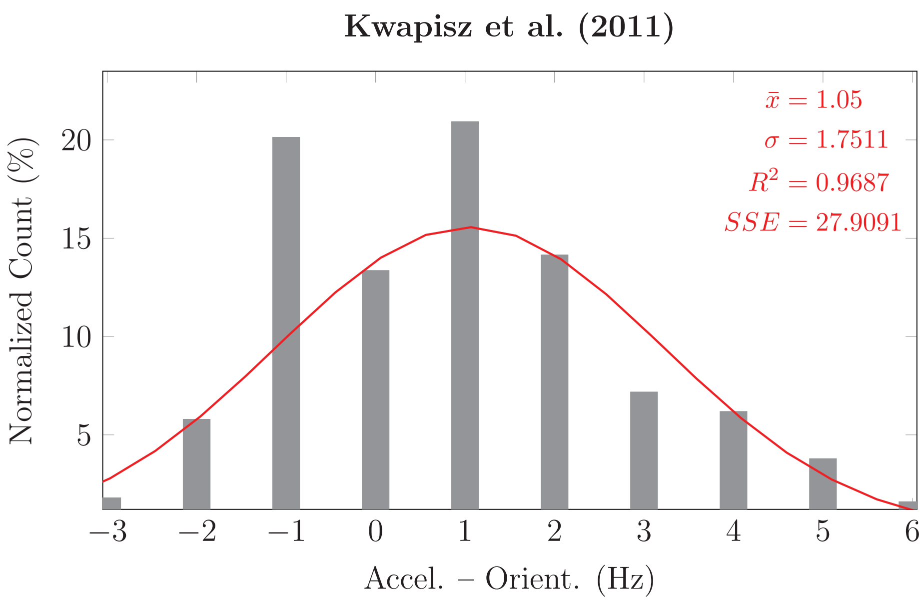


Figure 4.6: Distribution of differences in MES frequencies extracted from classification results of orientations subtracted from those extracted from accelerations for each activity and monitor. Results from using Bao and Intille's feature-set are shown on the upper graph while those of Kwapisz et al.'s feature-set are shown on the lower graph. A Gaussian curve has been fitted and the mean, standard deviation, R^2 and root mean squared error given.

MES frequencies extracted from success-rates of accelerations are on average higher than those extracted from success-rates of orientations for both Bao and Intille's feature-set (t=11.62, df=551, $p > 0.05$) and Kwapisz et al.'s feature-set (t=7.45, df=501, $p > 0.05$).

The Gaussian model fitted onto the results is found to have a R^2 values of 0.966 and 0.9687 for results obtained from Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively.

For Bao and Intille's feature-set, the average difference is 0.82Hz. While that of Kwapisz et al.'s feature-set is 1.05Hz .

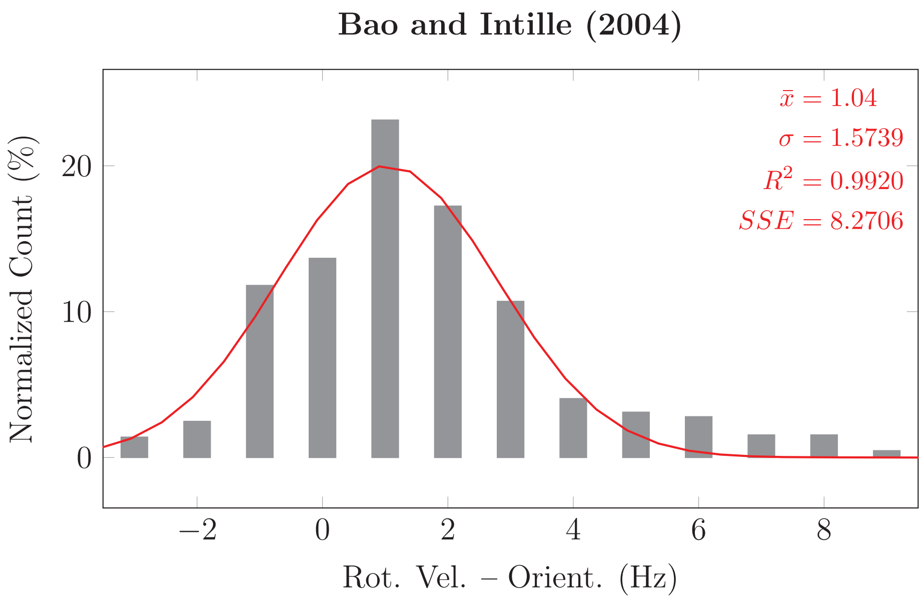
Do rotational velocity features require higher MES frequencies than orientation features?

The results of applying algorithm 3 on rotational velocities and orientations using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown in figure 4.7.

MES frequencies extracted from success-rates of rotational velocities are on average higher than those extracted from success-rates of orientations for both Bao and Intille's feature-set (t=11.52, df=643, $p > 0.05$) and Kwapisz et al.'s feature-set (t=19.99, df=734, $p > 0.05$).

The Gaussian model fitted onto the results is found to have a R^2 values of 0.992 and 0.9858 for results obtained from Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively.

For Bao and Intille's feature-set, the average difference is 1.04Hz. While that of Kwapisz et al.'s feature-set is 1.43Hz.



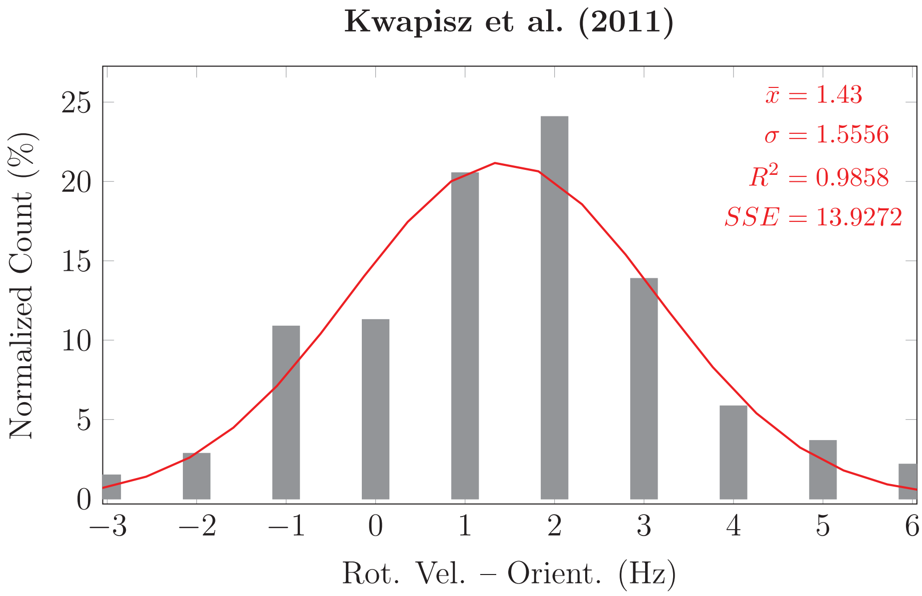


Figure 4.7: Distribution of differences in MES frequencies extracted from classification results of orientations subtracted from those extracted from rotational velocities for each activity and monitor. Results from using Bao and Intille's feature-set are shown on the upper graph while those of Kwapisz et al.'s feature-set are shown on the lower graph. A Gaussian curve has been fitted and the mean, standard deviation, R^2 and root mean squared error given.

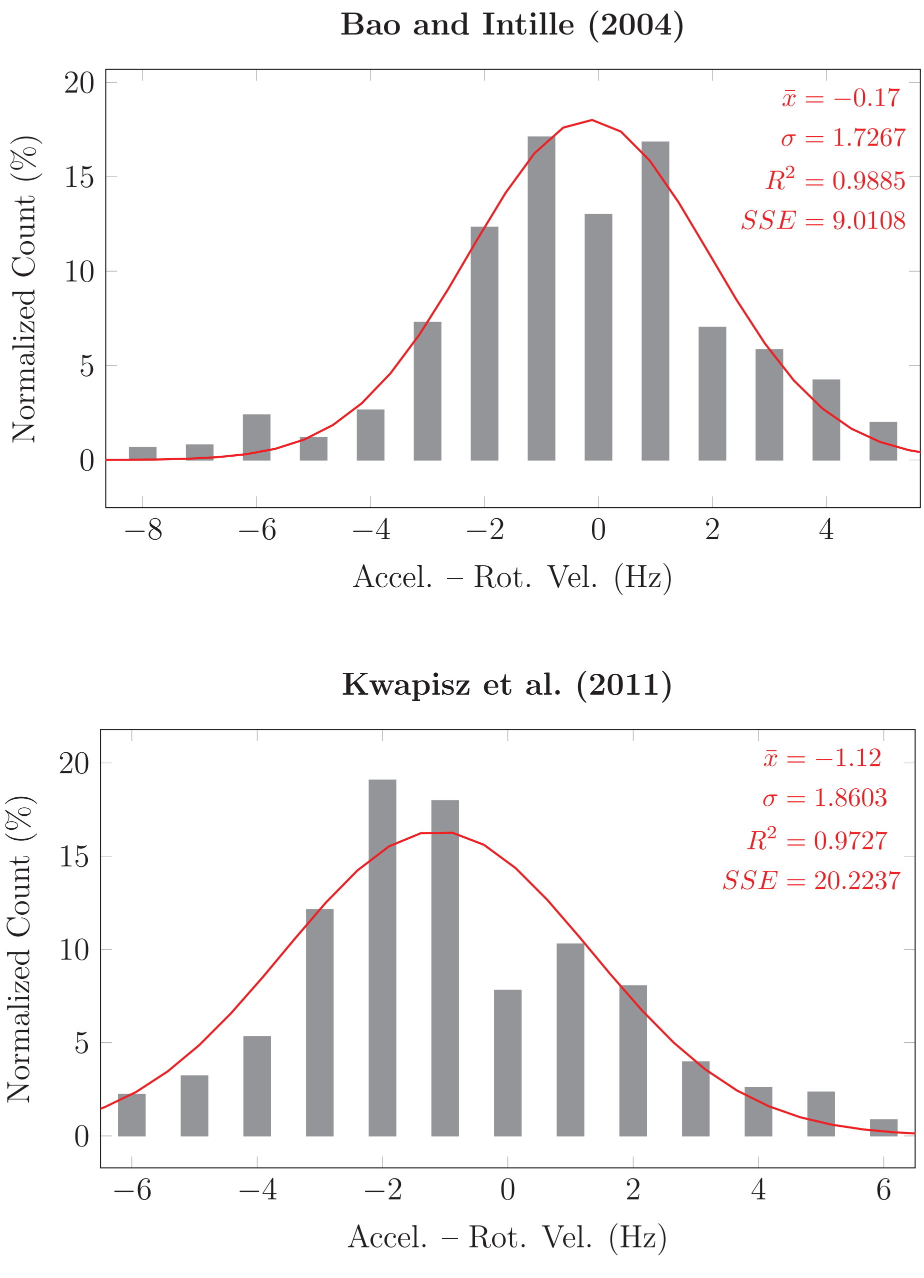


Figure 4.8: Distribution of differences in MES frequencies extracted from classification results of rotational velocities subtracted from those extracted from accelerations for each activity and monitor. Results from using Bao and Intille's feature-set are shown on the upper graph while those of Kwapisz et al.'s feature-set are shown on the lower graph. A Gaussian curve has been fitted and the mean, standard deviation, R^2 and root mean squared error given.

Do accelerometer features require higher MES frequencies than rotational velocity features?

The results of applying algorithm 3 on rotational velocities and orientations using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown in figure 4.8.

MES frequencies extracted from success-rates of accelerations are on average lower than those extracted from success-rates of rotational velocities for Bao and Intille's feature-set (t=-4.86, df=753, $p > 0.05$) and for Kwapisz et al.'s feature-set (t=-6.28, df=806, $p > 0.05$).

The Gaussian model fitted onto the results is found to have a R^2 values of 0.9885 and 0.9727 for results obtained from Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively.

For Bao and Intille's feature-set, the average difference is 0.17Hz. While that of Kwapisz et al.'s feature-set is 1.12Hz.

Discussion

The research question asked in this section was which *sources* require the highest and the lowest sampling frequencies. To answer this question, the MES frequencies of each activity and monitor were extracted as described by algorithm 2. The differences of MES frequencies from data obtained from different *sources* was then computed using algorithm 3.

The results showed that MES frequencies extracted from classification results of feature vectors extracted from orientations were lower than both those extracted from classification results of feature vectors extracted from accelerations and those extracted from classification results of feature vectors extracted from rotational velocities for the two feature-sets used.

Since MES frequencies represent the lowest frequency that is independent of the sampling frequency, lower MES frequencies extracted from classification results of features extracted from orientations imply that orientations require the lowest sampling frequency of the three *sources* studied and for the two feature-sets studied.

Additionally, the results showed that MES frequencies extracted from classification results of feature vectors extracted from rotational velocities were higher than both those extracted from classification results of feature vectors extracted from accelerations and those extracted from classification results of feature vectors extracted from orientations for the two feature-sets used.

Similarly, this implies that rotational velocities require the highest sampling frequency of the three *sources* studied and for the two feature-sets studied.

Even though the results obtained are statistically significant, the average differences between the three *sources* can be considered low. The largest average difference of 1.43Hz is between MES frequencies extracted from classification results of feature vectors extracted from rotational velocities and from classification results of feature vectors extracted from orientations using Kwapisz et al.'s feature-set. It is highly unlikely that a difference of 1.43Hz could impact battery consumption or computing resources.

4.3 Impact of sampling window length on activity classification accuracy

In this section, the impact of the length of the sampling window on classification success rates is studied. The length of the sampling window impacts the power consumption of the system since the longer the sampling window is the longer the sensor needs to operate and the more data is collected that requires processing.

In addition, long windows impact the response time of real-time activity recognition systems since sampling for a longer time means a longer duration between the time the first sample is taken and the time the activity classification is produced. This is clearly observed as a lag in the system recognising changes in activity when the user changes his/her activity.

However, a small window is likely to lead to a lower activity classification accuracy since the window is sampling characteristics of the activity motions and sampling error increases with decrease in sample size.

It is also possible that a window that is too long (in addition to impacting the power consumption) could impact activity classification accuracy by aggregating periods of the sensor signals that have different trends (perhaps due to the subject changing activities or changing how he/she is performing the activity).

In this section, we are interested in answering the following research questions:

1. **What is the relationship between activity classification success-rates and the length of the windows used to extract features for classification?**

   In this question, we are interested to learn what relationship exists between activity classification success-rates and window length. Intuitively the success-rates should increase with increasing window length, however, it is unclear whether this increase is linear or asymptotic.

   A better understanding of this relationship could provide activity recognition system developers a better understanding in the selection of an appropriate window length that takes account of factors like required system accuracy, response time and power consumption.

2. **Is there a window length above which activity classification success-rates are not affected by the length of the windows used to extract features for classification?**

   It is interesting to learn whether or not a limit exists in the relationship between activity classification success-rates and window length.

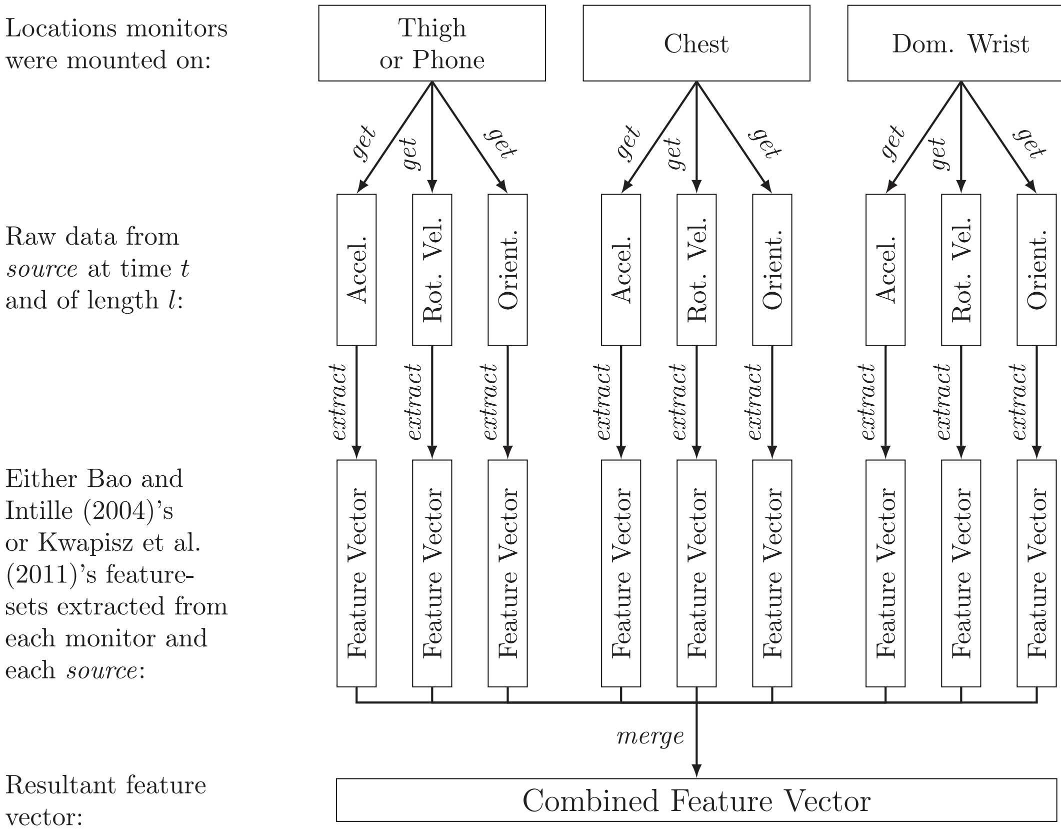


Figure 4.9: Overview of process taken to extract feature vector from time t and length l.

> If a system could sample indefinitely, is there a period after which further sampling has no impact on the success-rate of the system?

Methodology

Features were extracted from accelerations, rotational-velocities and orientations. The monitors mounted on the thigh, chest and dominant wrist were used. This allows all data from all activities (including running and walking) to be included.

The features of all the three *sources* from all three monitors were then combined such that the feature-vectors extracted from accelerations, rotational-velocities and orientations at time t and length l were merged together to form one feature-vector. Figure 4.9 illustrates the feature extraction process.

As observed in section 4.2, changes in downsampling frequency can cause fluctuations in the success-rate. Hence, the procedure was repeated for sampling frequencies in the range [121Hz,128Hz] in steps of 1Hz. The confusion matrixes were then averaged resulting in a mean confusion matrix. The range [121Hz,128Hz] was selected because an analysis of the results obtained in section 4.2 showed that success-rates obtained from downsampled frequencies in this range are generally constant although they showed minor fluctuations for all activities.

The procedure was repeated for window sizes of 1 second to 20 seconds in steps of one second. The windows were shifted by a 1 second period between one window and the next.

Algorithm 4 further elaborates this procedure.

Features were then extracted as explained in the respective feature-set's paper (Bao and Intille (2004) and Kwapisz et al. (2011), refer to section 3.5 for more details). In addition, a Hamming window was applied to each window before extracting the frequency-domain features in Bao and Intille's feature set.

Results

The results obtained in algorithm 4 are divided into sets. Each result set is obtained from classifying one fold, using 10-fold cross-validation, of feature-vectors extracted using either one of the studied feature-sets and using a single window length in the range of [1,20] seconds.

A one-sample two-tailed t-test of the activity classification result sets found each of the classification result sets to be significantly higher than chance (df=79, $p > .05$).

The result sets as a function of window length are presented in figure 4.10 for the two feature-sets studied.

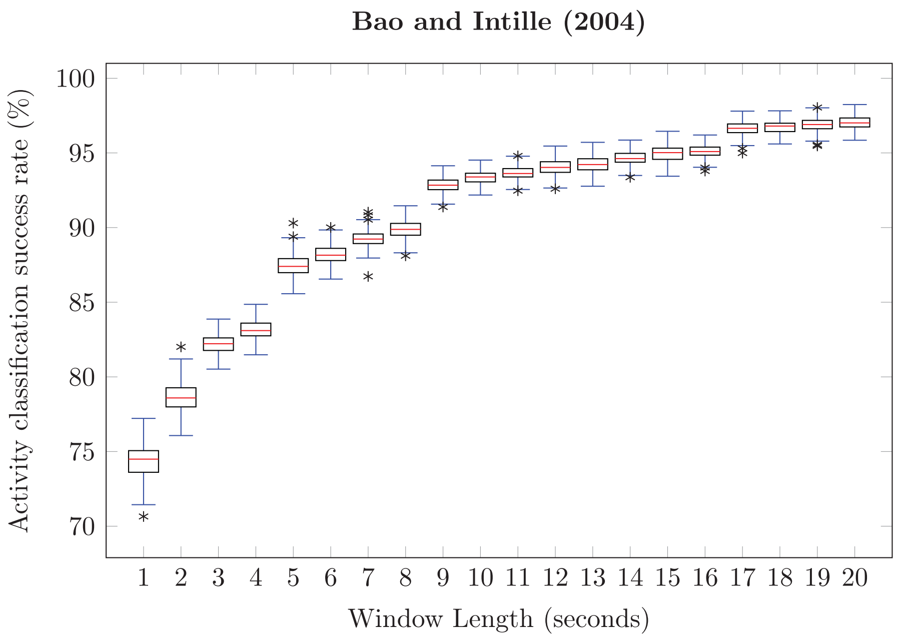
Algorithm 4 Procedure of testing the impact of window lengths on success-rates.

```
procedure ImpactOfWindowLength
    for all L ∈ [1s, 20s] in steps of 1s do
        ResultSet ← []                    ▷ Initialise empty list to hold obtained confusion matrixes.
        for all Fs ∈ [121Hz, 128Hz] in steps of 1Hz do
            D ← []                        ▷ Initialise feature vector data set.

            for all A ∈ Activities do
                for all S ∈ Subjects who did activity do
                    N ← duration of subject S doing activity A
                    for all t ∈ [0, N) in steps of 1s do
                        Extract feature vector V from data
                            of length L starting from time t
                            from source S of monitor M
                            on subject S doing activity A
                            downsampled to frequency Fs
                        Add V to D
                    end for
                end for
            end for

            F ← Split D into 10 folds
            for all f ∈ F do
                TrainingSet ← all folds except for f
                Train J48 classifier with TrainingSet
                Test with f
                Add success-rate to ResultSet
            end for
        end for

        Store ResultSet for further analysis.
    end for
end procedure
```



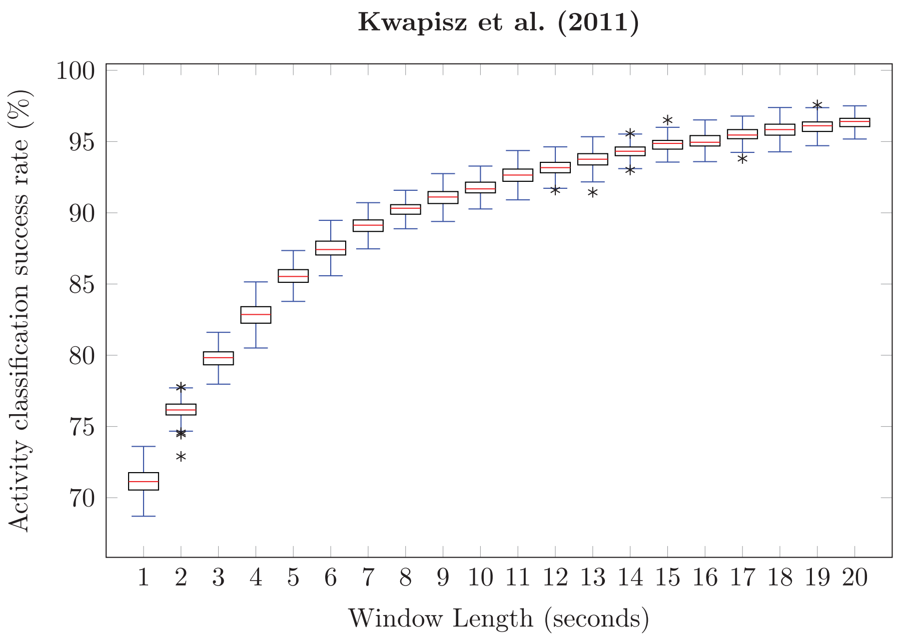


Figure 4.10: Activity classification success-rates obtained, presented as a function of window length for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al. (down).

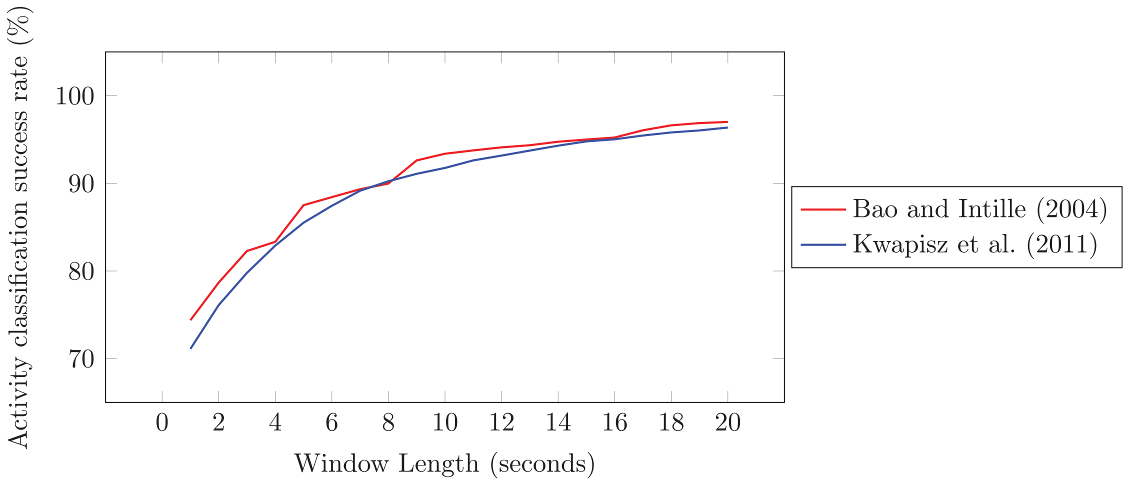


Figure 4.11: Mean success rates as obtained from algorithm 4 as a function of window length for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al. (down).

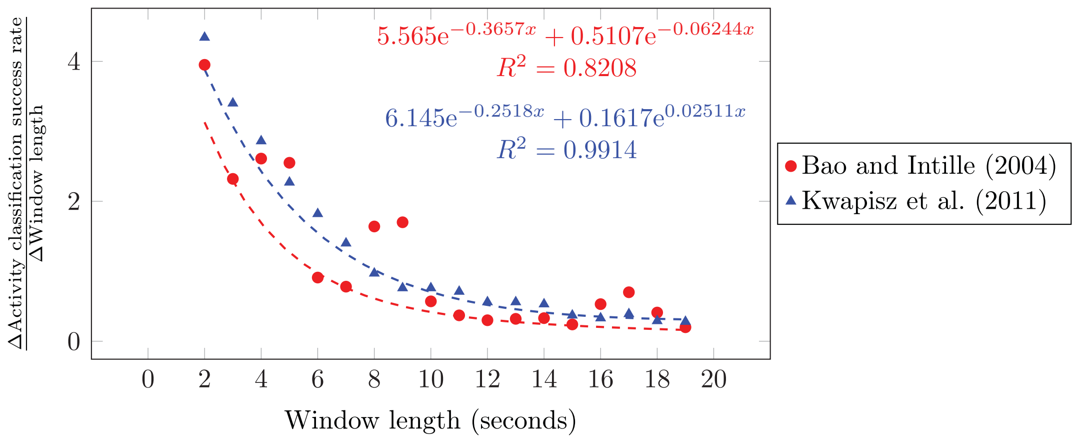


Figure 4.12: Derivative of mean success rates as obtained from algorithm 4 as a function of window length for the two studied feature-sets: Bao and Intille and Kwapisz et al. Sum of exponential equations have been fitted to show trends in the derivative of mean success-rates as functions of window length for the two studied feature-sets.

What is the relationship between success-rates and window length?

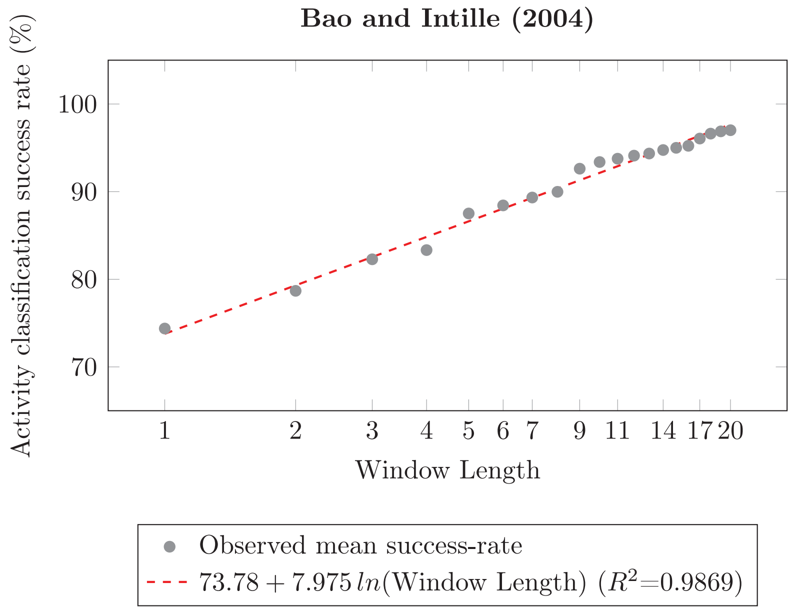
To answer this research question, the mean activity classification success-rates obtained from each result set as a function of window lengths were computed. The mean classification success-rates are presented in figure 4.11. Mean success-rates obtained from activity classification using feature-vectors extracted using Bao and Intille's feature-set as well as results from activity classification using feature vectors extracted using Kwapisz et al.'s feature-set are given.

From figure 4.11 we can observe that for both feature-sets studied, the mean classification success rates increase with increase in window length across the entire range of window lengths tested. However, the derivative of mean success rates decreases with increasing window-length for both feature-sets studied within the range of window lengths tested.

Figure 4.12 shows the derivative of mean success-rates as a function of window length for both feature-sets studied. From the figure, we can notice that the differential of the mean success rates is not linear with reference to window lengths. The change in success-rate is larger for smaller window lengths and smaller for larger window lengths. The trends observed in figure 4.12 suggest that the relationships between classification success-rates and window length are logarithmic for both feature-sets.

Figure 4.13 shows the mean success-rates plotted as a function of the natural logarithm of window length. From the figure, we can observe that the relationship between the success-rate and the logarithm of the window length is close to linear within the range of window lengths tested.

Fitting a linear model to the observed mean success-rates as a function logarithmic window lengths results in the linear models are given as equation 4.2 and equation 4.3 for Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively. The goodness of fit values of the two models are given in table 4.1.



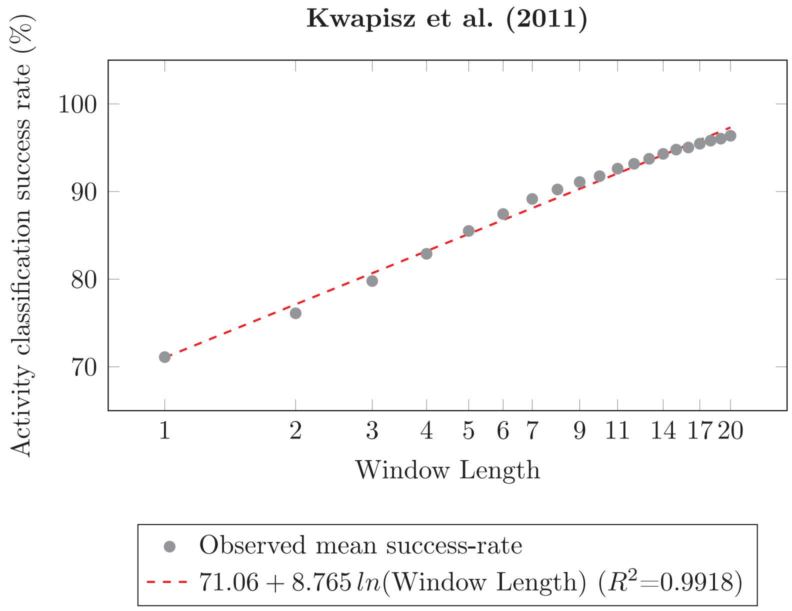


Figure 4.13: Mean success-rates of each result set obtained from algorithm 4 as a function of the natural logarithm of window lengths for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al (down). Fitted linear models of the data are also shown in the charts. Any logarithmic base results in a similar trend, however, the scale of the x-axis changes accordingly.

Table 4.1: Goodness of fit values for linear model fitted on the mean success-rates as a function of logarithmic window length for the two feature-sets studied.

Goodness of fit value	Bao and Intille (2004)	Kwapisz et al. (2011)
Sum of Square Error	10.63	7.945
R^2	0.9869	0.9918
Adjusted R^2	0.9861	0.9914
Root Mean Square Error	0.7683	0.6644

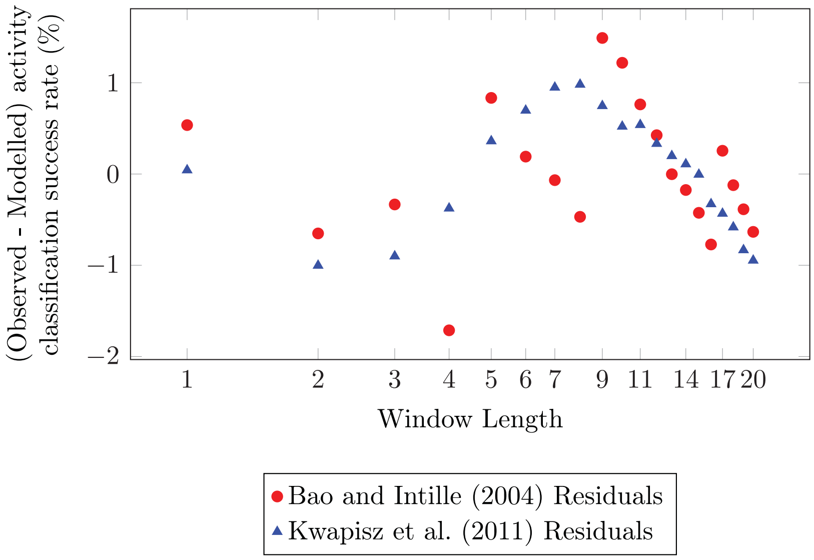


Figure 4.14: Residuals of fitting a linear model onto the mean success-rates as a function of the natural logarithm of window lengths as shown in figure 4.13 for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al (down).

$$\text{Success-rate}_{bao} \approx 73.78 + 7.975 ln(\text{Window Length}), \quad (4.2)$$

$$\text{Success-rate}_{kwapisz} \approx 71.06 + 8.765 ln(\text{Window Length}), \quad (4.3)$$

$$\text{where, Window Length} \in [1, 20]$$

The residuals of fitting linear models onto the data are shown in figure 4.14.

From the residuals scatter plot, we can observe that the residual obtained after fitting the linear model on mean success-rate results obtained from Kwapisz et al.'s feature-set have a sinusoidal pattern. This suggests we can further improve the model fitted on Kwapisz et al.'s feature-set result sets to reduce the difference between the model and the observed data by including sine curves into the model.

However, the sinusoidal pattern is not observed in residuals obtained after fitting the linear model on mean success-rate results obtained from Bao and Intille's feature-set. This indicates that the sinusoidal pattern observed might not be generalisable to the results obtained from other feature-sets.

Hence, within the range of window lengths tested, the activity classification success-rates obtained can be estimated as a linear function of logarithmic length of windows used to extract feature-vectors for classification. In other words, a classic asymptotic or diminishing-returns function.

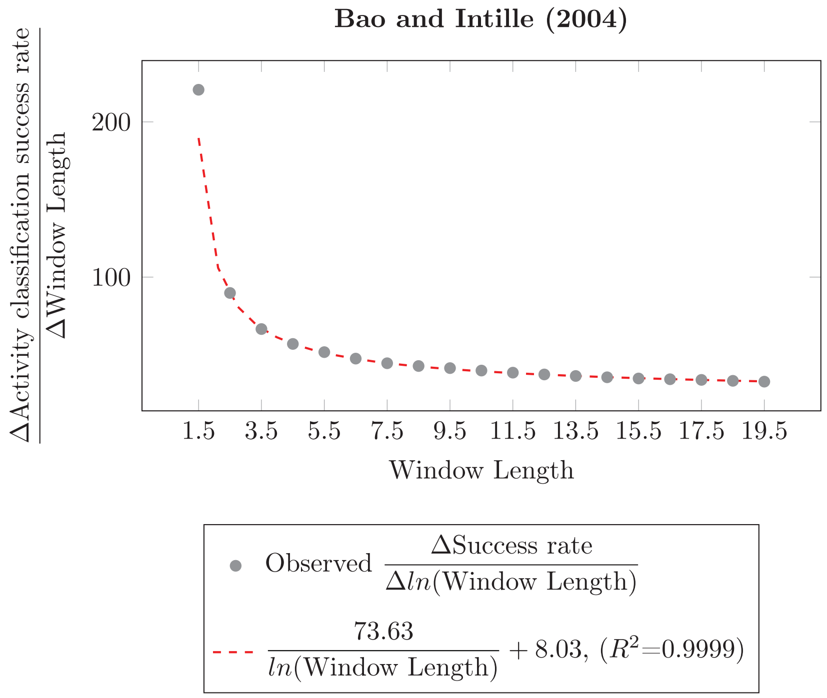
Is there a window length above which success-rates are not affected by window length?

To address this research question, the relationships estimated in the previous research question need to be extrapolated past the range of window lengths tested. To do this, it is important that we analyse the change in mean success-rate with change in logarithmic window length (i.e. the derivative of mean success-rate with respect to logarithmic window length).

The analysis presented in this only theoretical and purely based on the relationships observed so far. In practice, other factors that have not been considered in this analysis might impact the activity classification success-rates obtained for window lengths outside the tested range.

The plots of the derivative of mean success-rate with respect to logarithmic window length are given in 4.15 for both Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

In the figures, we can observe the decay of the derivative of mean success-rate with respect to logarithmic window length for both feature-sets. We can fit equa-



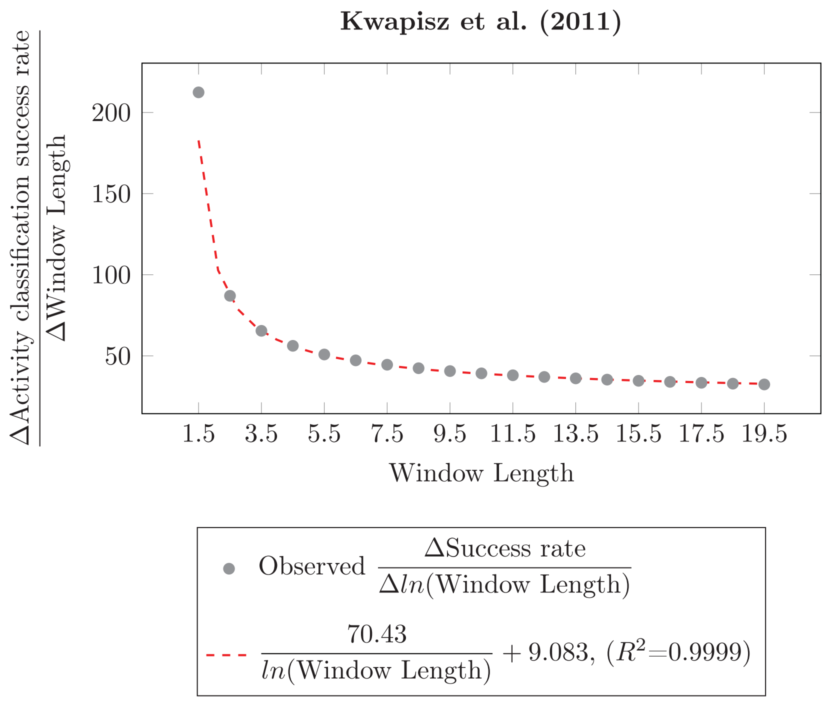


Figure 4.15: Derivative of mean success-rates of each result set obtained from algorithm 4 as a function of window length for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al (down). Two equations have been fitted so as to show the trend in the obtained mean success-rates as a function of the natural logarithm of window lengths.

tions composed of a constant term and a term that is inversely proportional to the logarithm of the window length to the data. The equation is given as 4.4.

$$\frac{\Delta SuccessRate}{\Delta ln(WindowLength)} \approx \frac{A}{ln(\text{Window Length})} + B, \tag{4.4}$$

$$\text{where } A \text{ and } B \text{ are constants}$$

$$\text{and Window Length } \in (1, 20]$$

After finding the best values for A and B using Non-linear Least Squares method, equation 4.5 is found to approximate the derivative of success-rates with respect to logarithmic window length for Bao and Intille's feature-set, while equation 4.6 is found to approximate the derivative of success-rates with respect to logarithmic window length for Kwapisz et al.'s feature-set. Equation 4.5 and equation 4.6 are plotted in figure 4.15.

$$\frac{\Delta SuccessRate_{bao}}{\Delta ln(WindowLength)} \approx \frac{73.63}{ln(\text{Window Length})} + 8.03, \tag{4.5}$$

$$\frac{\Delta SuccessRate_{kwapisz}}{\Delta ln(WindowLength)} \approx \frac{70.43}{ln(\text{Window Length})} + 9.083, \tag{4.6}$$

By extrapolating the two equations past the tested range of window lengths, we can observe that the limit of equation 4.5 is the constant term 8.03 as the window length approaches infinity. Similarly, the limit of equation 4.6 is the constant 9.083 as the the window length approaches infinity.

Hence, the two equations are always greater than zero. This indicates that the derivative of the mean success-rate as a function of logarithmic window will always be greater than zero but will continue to decrease as the window length increases.

However, due to the constant term, the mean success-rate is likely to increase till 100%, at which point, no further increase in window length would result to any further increase in success-rate. This is based on the model observed. The model

Table 4.2: Goodness of fit values for the refined model of the mean success-rates as a function of logarithmic window length for the two feature-sets studied.

Goodness of fit value	Bao and Intille (2004)	Kwapisz et al. (2011)
Sum of Square Error	8.771	2.836
R^2	0.9834	0.9954
Adjusted R^2	0.9813	0.9948
Root Mean Square Error	0.7404	0.421

observed is not 100% accurate and hence the window length obtained is highly theoretical.

To estimate the window length at which the mean success-rates intersect with 100%, the models estimated in the previous section need to be refined to include the model observed to fit the derivative. The models estimated in the previous section were linear and hence only reflected the constant term observed in the derivative.

The integral of equation 4.4 takes the form:

$$A\,ln(x) + Bx + C, \quad x = ln(\text{Window Length}) \tag{4.7}$$

Non-linear Least Squares method was used again to find the values of A, B and C for the two feature-sets studied. The final models of success-rates as a function of window lengths are presented as equation 4.8 for Bao and Intille's feature-set and equation 4.9 for Kwapisz et al.'s feature-set. Goodness of fit values are shown in table 4.2.

$$\text{Success-rate}_{bao} \approx 3.461\,ln(x) + 6.161x + 75.22 \tag{4.8}$$

$$\text{Success-rate}_{kwapisz} \approx 6.696\,ln(x) + 4.969x + 74.54 \tag{4.9}$$

$$\text{where, } x = ln(\text{Window Length})$$

$$\text{Window Length} \in [2, 20]$$

Using the refined models and extrapolating them past the range of window lengths tested, a success-rate of 100% is estimated to be achievable using window

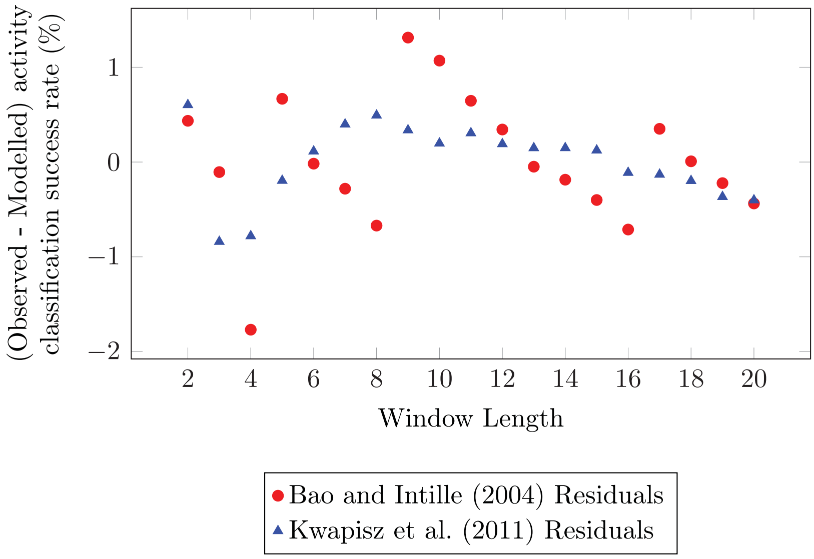


Figure 4.16: Residuals of fitting the refined models given in equation 4.8 and equation 4.9 onto the mean success-rates as a function of the natural logarithm of window lengths for the two studied feature-sets: Bao and Intille (up) and Kwapisz et al (down).

lengths of 28.33 seconds or higher using Bao and Intille's feature-set, and 31.61 seconds or higher using Kwapisz et al.'s feature-set. Figure 4.17 illustrates the estimated relationships between success-rate and window length when extrapolated to intersect with 100% success.

Discussion

In this section, the impact of window length on activity classification success-rates has been studied. Activity classification has been performed on all gathered activities using two different feature-sets using window lengths of 1 second to 20 seconds in increments of 1 second.

A linear relationship was found to exist between the mean success-rates obtained and the logarithm of the window length used to extract features, for both feature-sets, within the range of window lengths tested.

This relationship between success-rates obtained and the window length used to extract the features can be explained in terms of information captured within each

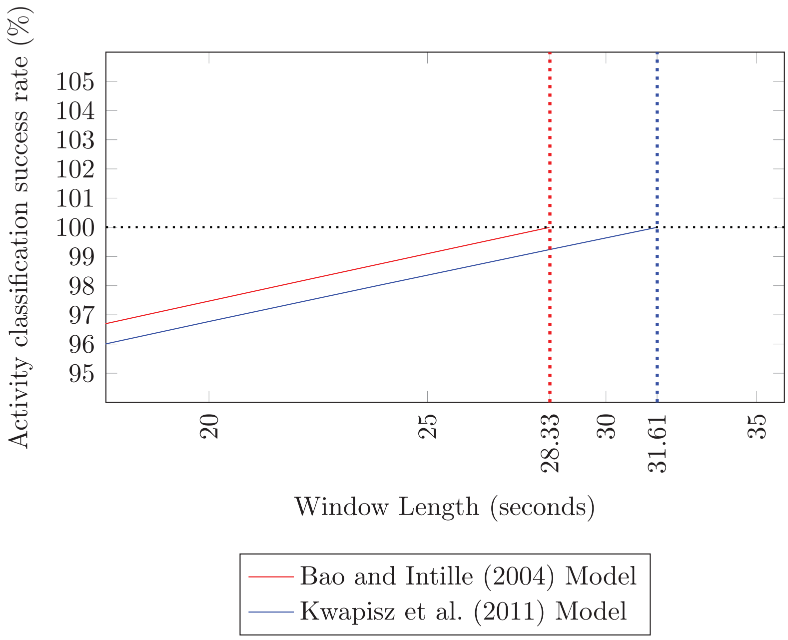


Figure 4.17: The estimated models between classification success-rate and window length for success-rates obtained using feature-vectors extracted using Bao and Intille's feature-set and Kwapisz et al.'s feature-set, extrapolated to intersect with 100% success-rate. The intersecting points are estimated to be 28.33 seconds for Bao and Intille's feature-set's relationship, and 31.61 seconds for Kwapisz et al.'s feature-set's relationship.

window.

Body motions display randomness that results in differences in similar short repeated motions (e.g. between one step and the next while walking). These random motions add variation to the motion signatures and hence add variation to the distributions of the feature-sets used to characterise the motions. This source of variability in the data can be termed as inter-repetition variability. Statistics calculated from each window incur sampling error which decreases with increase in the length of the window.

However, as the window length increases, the likelihood of encountering new information that is not already encompassed within the window grows lesser with increase in window length. Hence, larger increments of window length are required to achieve similar increments in success-rates at higher success-rates than at lower success-rates. This is a diminishing returns relationship between success-rates and window length.

The diminishing returns relationship is also evidenced by the model observed to fit the derivative of mean success-rates with respect to the logarithm of the window length. The model of the derivative is observed to contain a constant term and a term that is inversely proportional to the window length.

Given a long enough duration, the sampling error would reduce to zero. When the model of success-rates to logarithmic window lengths was refined to include observations made about its derivative, and extrapolated past the range of window lengths tested, it was estimated to intersect with 100% success-rate at 28.33 seconds Bao and Intille's feature-set, and 31.61 seconds for Kwapisz et al.'s feature-set. These intersections are, of course, only theoretical.

While the research has shown that, within the range of tested window lengths, increasing window lengths results in an increase in activity classification success-rate, other factors also have to be considered while selecting an appropriate window length. These factors include, but are not limited to:

1. The response rate required by the activity classification system. Increasing the window length increases the success-rate, but also means a longer duration is taken between starting sampling and producing the activity classification as well as in determining when one activity transitions to another.
2. Power consumption and processing requirements increase with increase in window length. Sampling longer windows means the sensor operates for a longer duration. It also means more data needs processing. The additional data processing consumes more power, CPU time and memory. All of which might be critical in a wearable system.

4.4 Impact of sampling window overlap on activity classification accuracy

In this section, the impact of the window overlap, or the window shift, on the success rate is studied. A larger overlap between sampling windows impacts the power consumption of the system since this leads to more windows for feature-extraction.

However, a larger overlap between windows also leads to more data to train and test the classifier with. It is unclear whether the extra data improves classification or not.

We are interested in answering the following research questions:

1. **Does the overlap size of the windows used to extract features for classification impact activity classification success-rates?**

   In this question, we are interested to learn whether or not window overlaps have any impact on activity classification success-rates. It is unclear whether changes in window overlap impact activity classification success-rates or not.

2. **If window overlap size does impact activity classification success-rates, what is the relationship between activity classification success-rates and overlap size of the windows used to extract features for classification?**

   If window overlaps are found to impact success-rates, a better understanding of the relationship could provide activity classification researchers a better understanding while selecting the window overlap to use while developing models and training activity classifiers.

3. **Is there a relationship between variance in activity classification success-rates and window overlap size?**

   In this question, we are interested in learning whether window overlaps impact the variance in activity classification success-rates. If they do, we are interested in characterising the relationship between window overlap and resultant activity classification success-rates variance.

Methodology

Features were extracted from accelerations, rotational-velocities and orientations. The monitors mounted on the thigh, chest and dominant wrist were used. This allows all data from all activities (including running and walking) to be included.

The features of all the three *sources* from all three monitors were then combined such that the feature-vectors extracted from accelerations, rotational-velocities and orientations at time t and length l were merged together to form one feature-vector. The feature-vectors were merged as was shown in section 4.3 which was illustrated using figure 4.9.

As observed in section 4.2, changes in downsampling frequency impact the success-rate. Hence, the procedure was repeated for sampling frequencies in the range [121Hz,128Hz] in steps of 1Hz. The confusion matrixes were then averaged resulting in a mean confusion matrix. The range [121Hz,128Hz] was selected because an analysis of the results obtained in section 4.2 showed that success-rates obtained from downsampled frequencies in this range are generally constant although they fluctuated for all activities.

A window size of 10 seconds was used. The procedure was repeated for window shifts of 1 second to 10 seconds (i.e. 90% overlap down to 0% overlap) in intervals of 1 second.

Window lengths of 10 seconds were used because 10 seconds was observed to be the maximum window length used in activity recognition literature, having only been used by Kwapisz et al. (2011) and Patel et al. (2009). In the analysis of activity recognition literature performed by Lockhart and Weiss (Lockhart & Weiss, 2014), window lengths reported to have been used in activity recognition literature were observed to have a median of 3 seconds and the maximum window length they observed was 10 seconds.

In addition, in the analysis conducted in section 4.3, in the analysis of change of success-rate as a function of change of window length, it was observed that window lengths of 10 seconds are in the 'plateau region' of the relationship.

Features were then extracted as explained in the respective feature-set's paper (Bao and Intille (2004) and Kwapisz et al. (2011), refer to section 3.5 for more details). In addition, a Hamming window was applied to each window before extracting the frequency-domain features in Bao and Intille's feature set.

Both 10-fold cross-validation and remove-one-subject cross-validation were used.

Algorithm 5 further elaborates on this procedure.

Results

The results obtained using algorithm 5 are divided into sets. Each result set is obtained from classifying one fold, either using 10-fold cross-validation or remove-one-subject cross-validation, of feature-vectors extracted using either one of the studied feature-sets and using a single window overlap in the range [0%,90%].

A one-sample two-tailed t-test of the activity classification result sets found each of the classification result sets to be significantly higher than chance for 10-fold cross-validation (df=79, $p > .05$) as well as for remove-one-subject cross-validation (df=143, $p > .05$).

The result sets of 10-fold cross-validation as a function of window overlap are presented in figure 4.18 for the two feature-sets studied.

Similarly, the result sets of remove-one-subject cross-validation as a function of window overlap are presented in figure 4.19 for the two feature-sets studied.

Algorithm 5 Procedure of testing the impact of window overlaps on success-rates.

```
procedure ImpactOfWindowOverlaps
    for all Overlap ∈ [0%,90%] in steps of 10% do
        TenFoldResultSet ← []      ▷ Initialise empty list to hold obtained confusion matrixes from 10-fold cross-validation.
        Rem1SubResultSet ← []      ▷ Initialise empty list to hold obtained confusion matrixes from remove-one-subject cross-validation.
        for all Fs ∈ [121Hz, 128Hz] in steps of 1Hz do
            D ← []                 ▷ Initialise feature vector data set.

            for all A ∈ Activities do
                for all S ∈ Subjects who did activity do
                    N ← duration of subject S doing activity A
                    for all t ∈ [0, N) in steps of (1 − Overlap/100) × 10 seconds do
                        Extract feature vector V from data
                            of length 10 seconds starting from time t
                            from source S of monitor M
                            on subject S doing activity A
                            downsampled to frequency Fs
                        Add V to D
                    end for
                end for
            end for
                                   ▷ Perform 10-fold cross-validation on feature-vectors in D.
            F ← Split D into 10 folds
            for all f ∈ F do
                TrainingSet ← all folds except for f
                Train J48 classifier with TrainingSet
                Test with f
                Add success-rate to TenFoldResultSet
            end for
                                   ▷ Perform remove-one-subject cross-validation on feature-vectors in D.
            F ← Split D into subject-based folds
            for all f ∈ F do
                TrainingSet ← all folds except for f
                Train J48 classifier with TrainingSet
                Test with f
                Add success-rate to Rem1SubResultSet
            end for
        end for

        Store TenFoldResultSet for further analysis.
        Store Rem1SubResultSet for further analysis.
    end for
end procedure
```

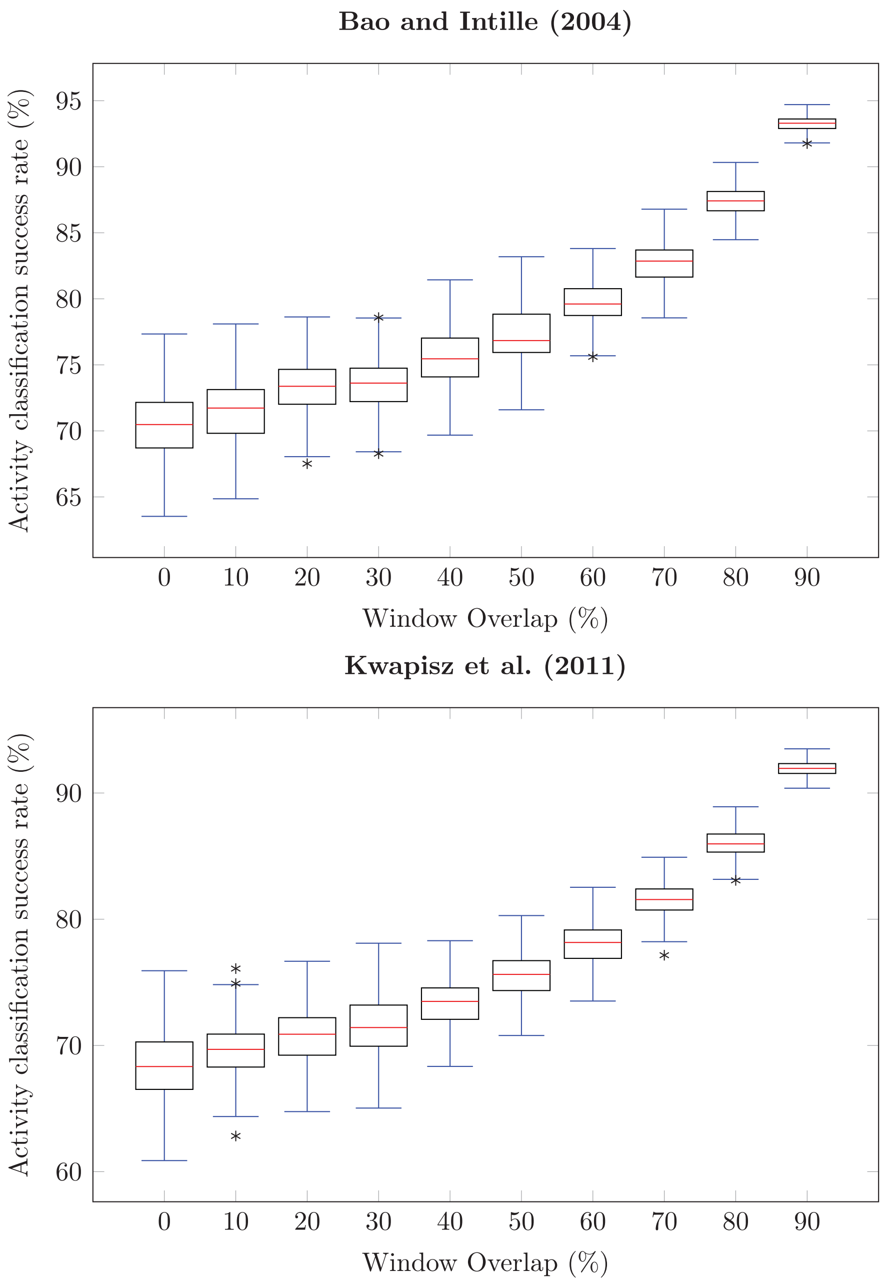


Figure 4.18: 10-fold cross-validation activity classification success-rates obtained, presented as a function of window overlap for the two studied feature-sets: Bao and Intille (upper) and Kwapisz et al. (lower).

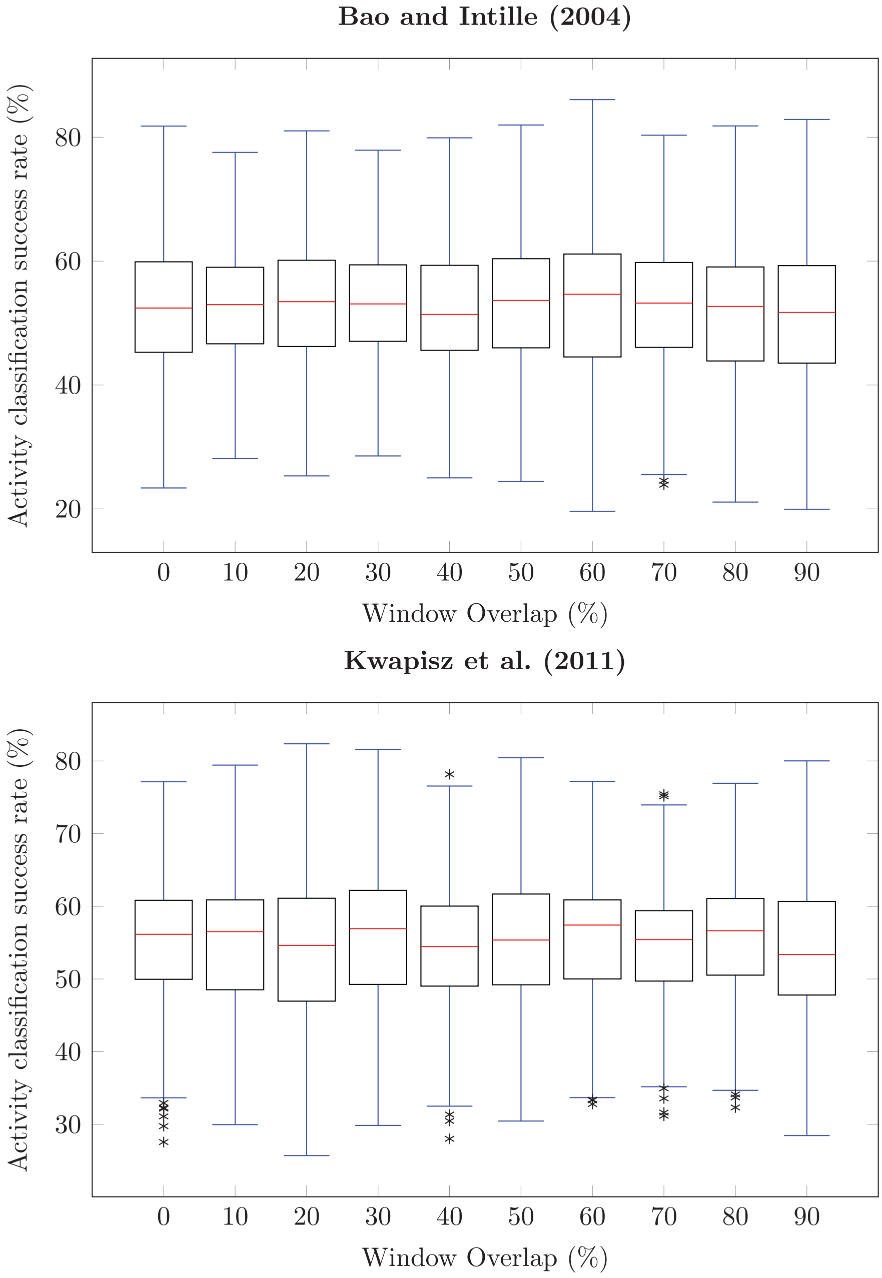


Figure 4.19: Remove-one-subject cross-validation activity classification success-rates obtained, presented as a function of window overlap for the two studied feature-sets: Bao and Intille (upper) and Kwapisz et al. (lower).

Does the window overlap size used to extract features for classification impact activity classification success-rates?

A two-tailed two-sample T-Test with $\alpha = 0.05$ between 10-fold cross-validation success-rates obtained using a window overlap of 0% and those obtained using a window overlap of 90% rejected the null hypothesis that the two groups of success-rates were sampled from normal distributions of equal means but unknown variances (t=-86.04, df=158, $p > 0.05$).

Similar tests were run between all possible pairs of 10-fold cross-validation result sets obtained using Bao and Intille's feature-set and between all possible pairs of 10-fold cross-validation result sets obtained using Kwapisz et al.'s feature-set. All the tests except for one rejected the null hypothesis that the result sets were random samples drawn from normal distributions of equal means and unknown variance. The test that rejected the null hypothesis was between success-rates obtained using a window overlap of 20% and those obtained using a window overlap of 30% using Bao and Intille's feature-set.

This means that all except one of the result sets obtained using 10-fold cross-validation from the same feature-set but different window overlaps are not statistically similar, and hence shows that window overlaps have an impact on activity classification success-rates in almost all of the cases.

Two-tailed two-sample T-Tests with $\alpha = 0.05$ were also run between all possible pairs of remove-one-subject cross-validation result sets obtained using Bao and Intille's feature-set and between all possible pairs of remove-one-subject cross-validation result sets obtained using Kwapisz et al.'s feature-set. All the tests except for one failed to reject the null hypothesis that the result sets were random samples drawn from normal distributions of equal means and unknown variance. The test that failed to reject the null hypothesis was between success-rates obtained using a window overlap of 30% and those obtained using a window overlap of 90% using

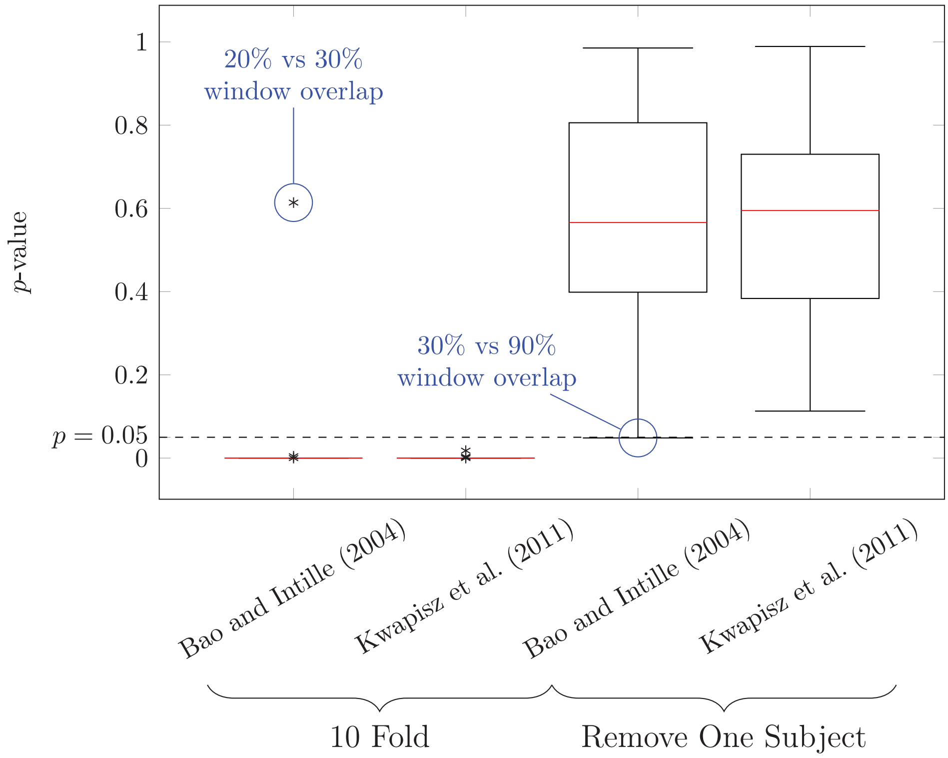


Figure 4.20: Box plot of p-values obtained from two-tailed two-sample t-Tests between all possible pairs of result sets obtained using algorithm 5 for the two studied feature-sets (Bao and Intille's and Kwapisz et al.'s feature-sets) and for 10-fold and remove-one-subject cross-validation. The two-tailed two-sample t-Tests test the null hypothesis that pairs of result sets obtained from the same feature-set and tested using the same cross-validation method but differing in window overlap come from independent random samples from normal distributions with equal means and equal but unknown variances. At the preselected significance level of 0.05, a p-value of 0.05 or below results in the rejection of the null hypothesis, while a p-value above 0.05 results in the failure to reject the null hypothesis.

Bao and Intille's feature-set.

This means that, while there is a chance that window overlaps do impact success-rates while using remove-one-subject cross-validation, the likelihood is low in almost all cases.

The two-tailed two-sample T-Test results are summarised in figure 4.20.

If window overlap size does impact activity classification success-rates, what is the relationship between activity classification success-rates and overlap size of the windows used to extract features for classification?

To answer this question, the analysis will focus on results of 10-fold cross-validation, since the chance of window overlaps impacting results of remove-one-subject cross-validation were found to be extremely low in the previous research question.

Figure 4.21 shows the mean success-rate as a function of window overlap using 10-fold and remove-one-subject cross-validation for the two studied feature-sets: Bao and Intille and Kwapisz et al.

For 10-fold cross-validation a clear pattern can be observed while no apparent pattern is observable in the remove-one-subject cross-validation mean success-rates.

The pattern observed in the 10-fold cross-validation results can be modelled as a power function with a constant offset in the form:

$$\text{mean success-rate} \approx A + B\,(\text{window overlap})^{C} \tag{4.10}$$

where A, B and C are constants.

Fitting the equation using Non-linear Least Squares method resulted in equation 4.11 and equation 4.12 for results obtained from Bao and Intille's feature-set and results obtained from Kwapisz et al.'s feature-set respectively. The goodness of fit values of the two equations to the observed success-rates are given in table 4.3 while the residuals are given in figure 4.22.

Fractional window overlaps are used instead of percentages in equation 4.11 and equation 4.12 because using percentages results in small B values.

$$\text{mean success-rate}_{bao} \approx 69.53 + 28.74\,(\text{fractional window overlap})^{2.309} \tag{4.11}$$

$$\text{mean success-rate}_{kwapisz} \approx 67.07 + 29.98\,(\text{fractional window overlap})^{2.258} \tag{4.12}$$

The residuals shown in figure 4.22 can be observed to be highly correlated. The computed correlation coefficient of the residuals obtained from the two studied

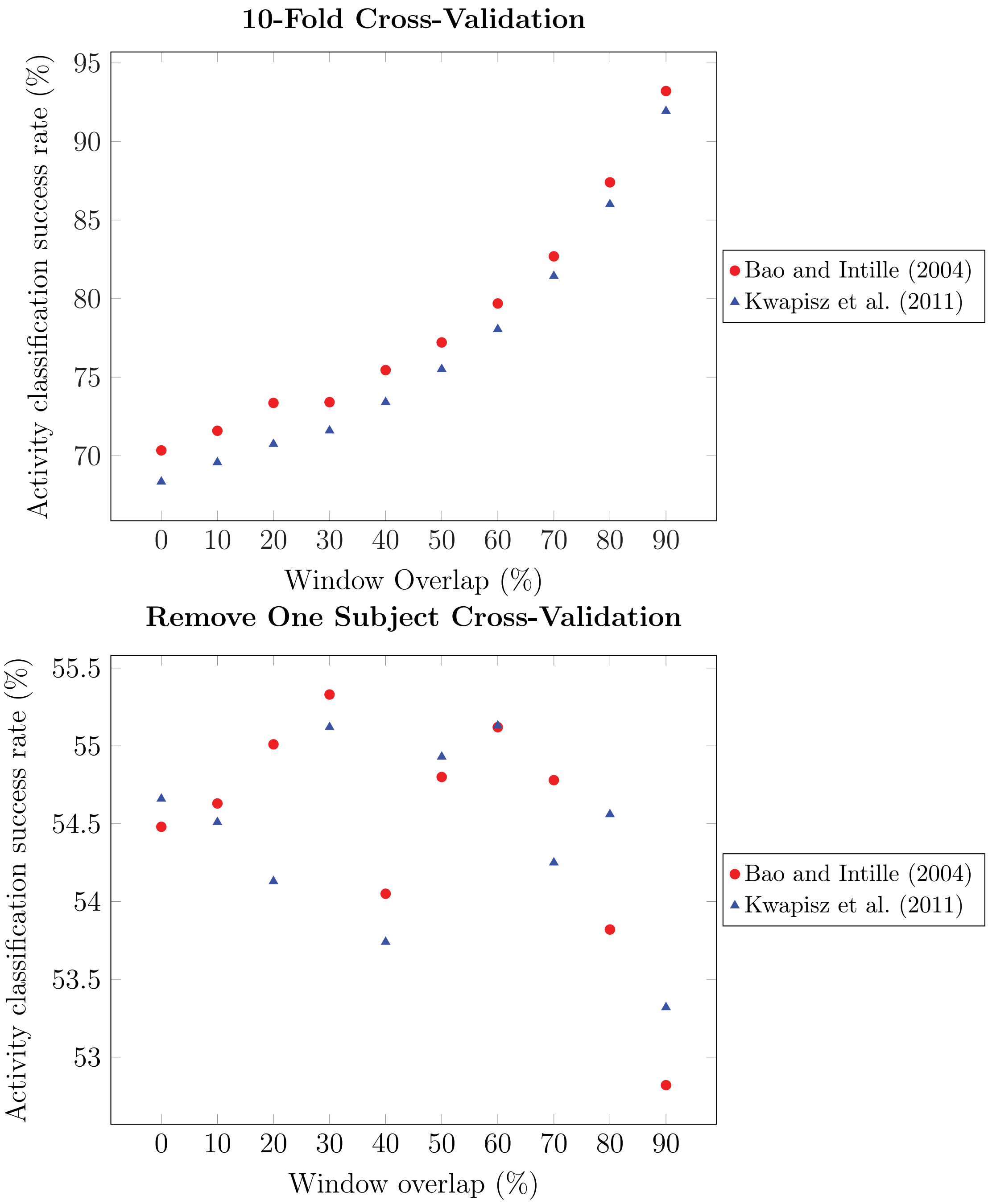


Figure 4.21: 10-fold (upper) and remove-one-subject (lower) cross-validation mean success-rates as computed using algorithm 5 as a function of percentage window overlap.

Table 4.3: Goodness of fit values for fitting equation 4.11 and equation 4.12 on mean success-rates as a function of window overlap for the two feature-sets studied.

Goodness of fit value	**Bao and Intille (2004)**	**Kwapisz et al. (2011)**
Sum of Square Error	4.321	2.412
R^2	0.9923	0.9961
Adjusted R^2	0.9901	0.995
Root Mean Square Error	0.7856	0.587

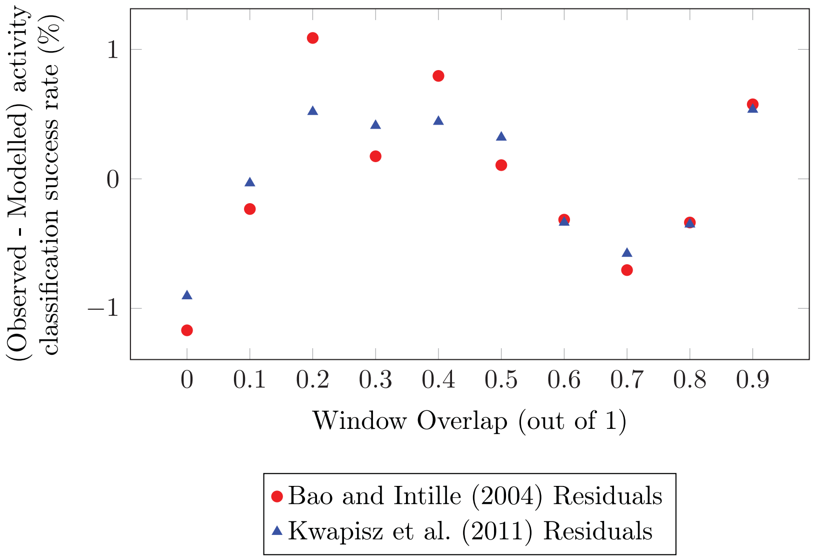


Figure 4.22: Residuals of fitting equation 4.11 and equation 4.12 on mean success-rates as a function of distance between windows for the two feature-sets studied.

feature-sets is 0.8912. This indicates that it is possible to further refine the models so as to cater for the observed common pattern between the residuals of the two feature-sets. However, from the data, it is unclear what equation can be used to model the residuals. In addition, the residuals are small (with a range of less than 2%), hence the error introduced to the model is also low.

From equation 4.11 and equation 4.12, we can say that the mean activity classification success-rates obtained using 10-fold cross-validation are proportional to window overlap raised to a constant power.

Alternatively, we can model the observed mean success-rate as a function of

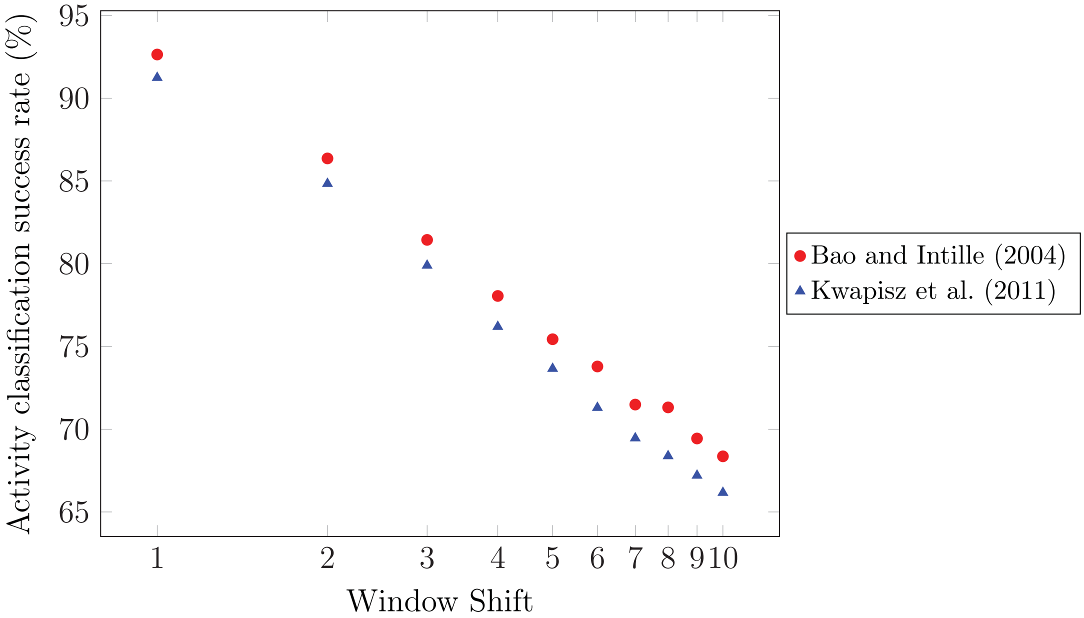


Figure 4.23: 10-fold cross-validation mean success-rates as computed using algorithm 5 as a function of logarithmic distance between windows for the two studied feature-sets Bao and Intille and Kwapisz et al.

logarithmic distance between windows. This approach worked well for section 4.3. 10-fold cross-validation mean success-rates as a function logarithmic distance between windows are shown in figure 4.23.

From figure 4.23, the mean success-rates as a function of logarithmic distance between windows are observed to approximate a linear relationship for both studied feature-sets. Equation 4.13 and equation 4.14 were found to be the best fitting linear equations with respect to the mean success-rates as a function of logarithmic distance between windows.

$$\text{mean success-rate}_{bao} \approx 93.06 - 10.74\, ln(\text{window shift}) \tag{4.13}$$

$$\text{mean success-rate}_{kwapisz} \approx 91.83 - 11.26\, ln(\text{window shift}) \tag{4.14}$$

Goodness of fit values of the two equations on the data are shown in table 4.4. Residuals of fitting the two equations on the data are shown in figure 4.24.

Similar to the residuals of equation 4.11 and equation 4.11 shown in figure 4.22, the residuals equation 4.13 and equation 4.14 shown in figure 4.24 are observed to

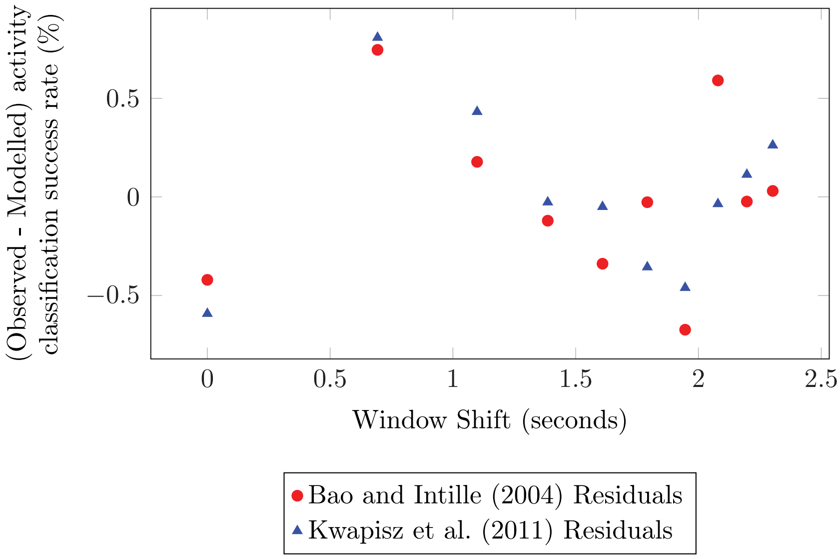


Figure 4.24: Residuals of fitting equation 4.13 and equation 4.14 on mean success-rates as a function of logarithmic distance between windows for the two feature-sets studied.

be correlated. The computed correlation coefficient of the residuals obtained from the two studied feature-sets is 0.7569.

Table 4.4: Goodness of fit values for fitting equation 4.13 and equation 4.14 on mean success-rates as a function of logarithmic distance between windows for the two feature-sets studied.

Goodness of fit value	**Bao and Intille (2004)**	**Kwapisz et al. (2011)**
Sum of Square Error	1.697	1.613
R^2	0.997	0.9974
Adjusted R^2	0.9966	0.997
Root Mean Square Error	0.4606	0.4491

Using the second set of equations (equation 4.13 and equation 4.14), we can say that a linear relationship with a negative slope exists between the mean activity classification success-rates obtained using 10-fold cross-validation and the logarithm of the distance between windows within the tested range. Hence, an increase in logarithmic window shift leads to a decrease in mean activity classification success-rates within the tested range.

From the goodness of fit values of the two sets of equations (compare table 4.3 to table 4.4) we can observe that the second set of equations fit the data slightly better. In addition, the second set requires one less parameter than the first set.

Is there a relationship between variance in activity classification success-rates and window overlap size?

Figure 4.25 shows the standard deviation of result set success-rates as a function of window overlap using 10-fold and remove-one-subject cross-validation for the two studied feature-sets: Bao and Intille and Kwapisz et al.

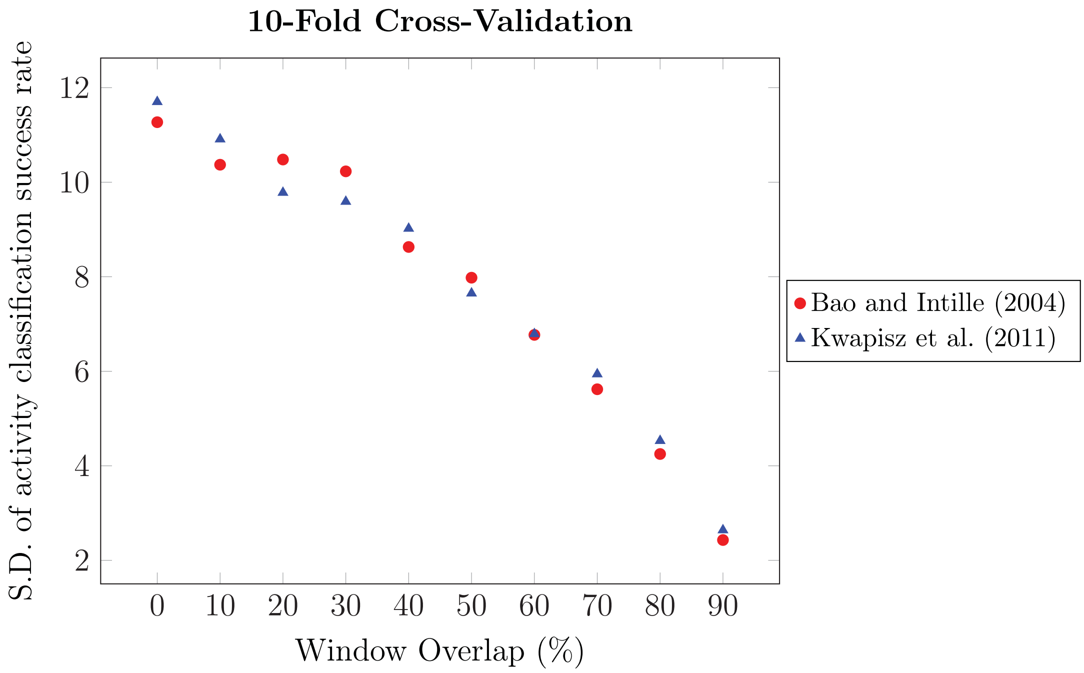
For 10-fold cross-validation a clear pattern can be observed while no apparent pattern is observable in the remove-one-subject cross-validation standard deviation of result set success-rates.

Two-tailed two-sample F-Tests with $\alpha = 0.05$ were run between all possible pairs of 10-fold cross-validation result sets obtained using Bao and Intille's feature-set and between all possible pairs of 10-fold cross-validation result sets obtained using Kwapisz et al.'s feature-set for both 10-fold and remove-one-subject cross-validation result sets. The results are summarised at figure 4.26.

In figure 4.26, we can observe that, for most of the result sets obtained using remove-one-subject cross-validation, the two-tailed two-sample F-Tests failed to reject the null hypothesis that result sets came from normal distributions of the same variance. However, for most of the result sets obtained using 10-fold cross-validation using Kwapisz et. al.'s feature-set and some of the result sets obtained using 10-fold cross-validation using Bao and Intille's feature-set, it rejected the null hypothesis.

Moreover, the medians of the distributions of p-values are seen to be below 0.05 for the 10-fold cross-validation result sets, and above 0.05 for the remove-one-subject cross-validation result sets.

Hence, we can say that the likelihood that window overlaps have an impact on the variance in remove-one-subject cross-validation activity classification results



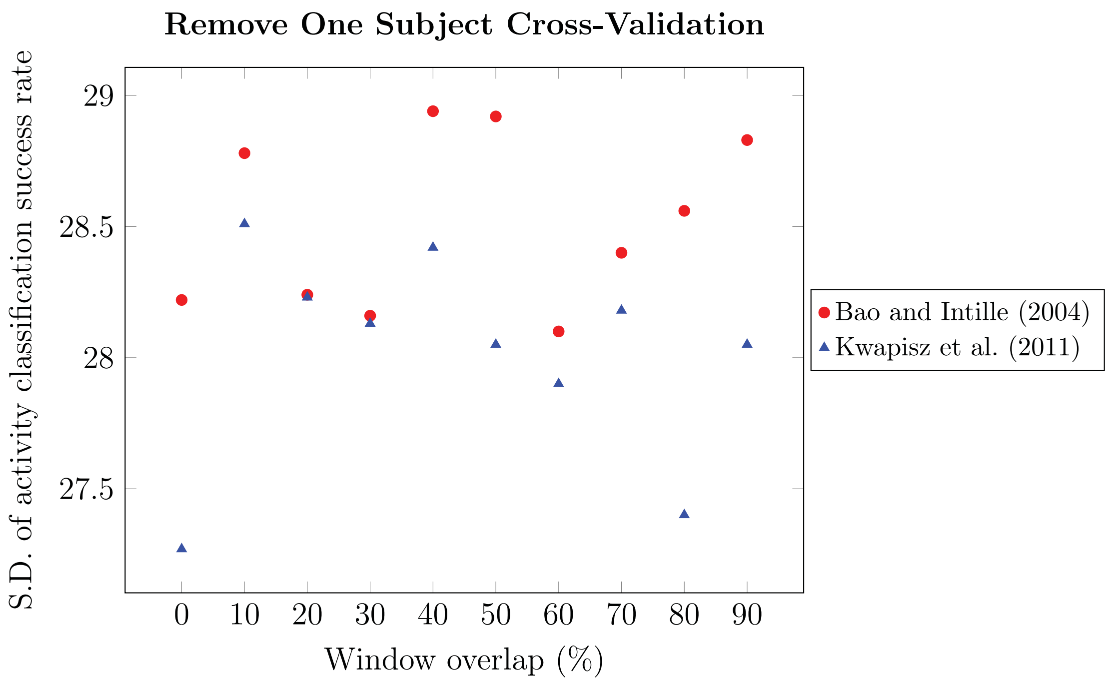


Figure 4.25: 10-fold (upper) and remove-one-subject (lower) cross-validation standard deviations of result-set success-rates as computed using algorithm 5 as a function of percentage window overlap.

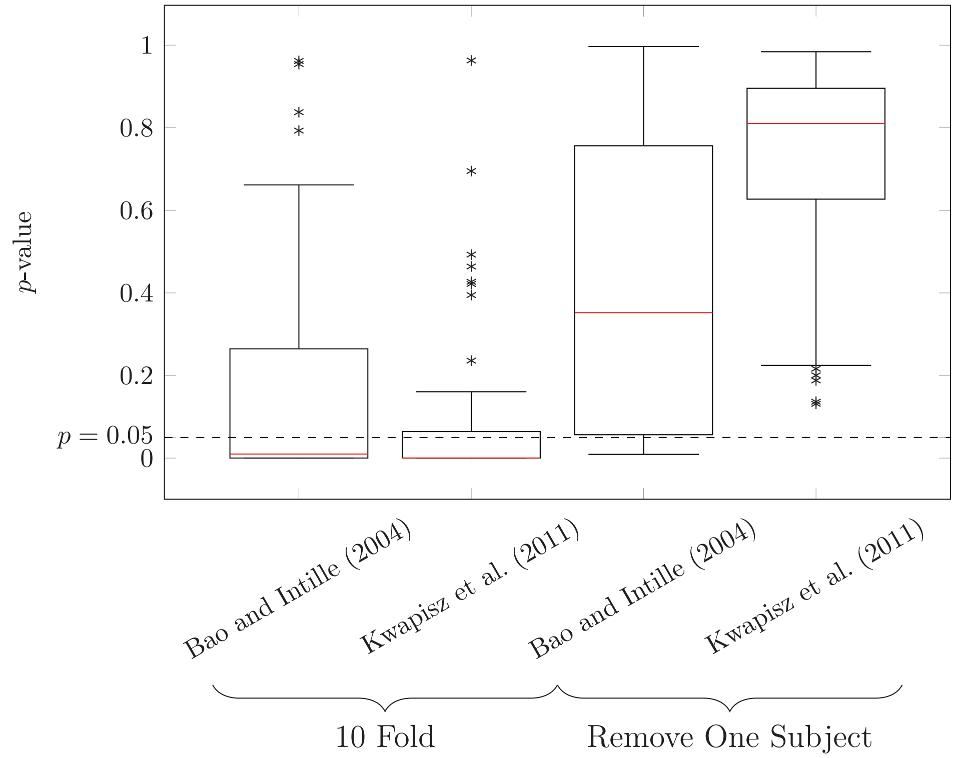


Figure 4.26: Box plot of p-values obtained from two-tailed two-sample f-Tests between all possible pairs of result sets obtained using algorithm 5 for the two studied feature-sets (Bao and Intille's and Kwapisz et al.'s feature-sets) and for 10-fold and remove-one-subject cross-validation. The two-tailed two-sample f-Tests test the null hypothesis that pairs of result sets obtained from the same feature-set and tested using the same cross-validation method but differing in window overlap come from normal distributions with the same variance. At the preselected significance level of 0.05, a p-value of 0.05 or below results in the rejection of the null hypothesis, while a p-value above 0.05 results in the failure to reject the null hypothesis.

is negligeable, while the likelihood that window overlaps have an impact on the variance in 10-fold cross-validation activity classification results is high.

As was observed in the previous research question, the mean success-rates could be modelled as a function of logarithmic window shift (the distance between windows in seconds). 10-fold cross-validation standard deviations of result set success-rates as a function logarithmic distance between windows are shown in figure 4.23.

From figure 4.23, the standard deviations of result set success-rates as a function of logarithmic distance between windows are observed to approximate a linear relationship for both studied feature-sets. Equation 4.15 and equation 4.16 were found

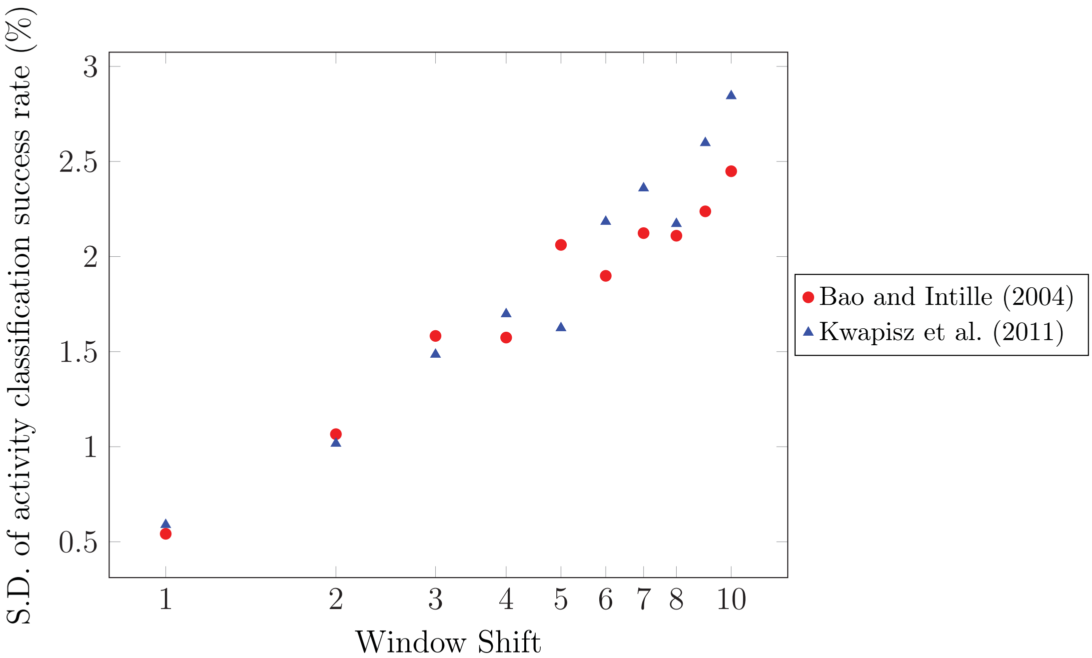


Figure 4.27: 10-fold cross-validation standard deviations of result set success-rates as computed using algorithm 5 as a function of logarithmic distance between windows for the two studied feature-sets Bao and Intille and Kwapisz et al.

to be the best fitting linear equations with respect to the mean success-rates as a function of logarithmic distance between windows.

$$\text{S.D. success-rate}_{bao} \approx 0.5766 + 0.7864\, ln(\text{window shift}) \tag{4.15}$$

$$\text{S.D. success-rate}_{kwapisz} \approx 0.4415 + 0.9372\, ln(\text{window shift}) \tag{4.16}$$

Goodness of fit values of the two equations on the data are shown in table 4.5. Residuals of fitting the two equations on the data are shown in figure 4.28.

Table 4.5: Goodness of fit values for equation 4.13 and equation 4.14 on the standard deviations of result set success-rates as a function of logarithmic distance between windows for the two feature-sets studied.

Goodness of fit value	**Bao and Intille (2004)**	**Kwapisz et al. (2011)**
Sum of Square Error	0.1076	0.2653
R^2	0.9653	0.9412
Adjusted R^2	0.9609	0.9339
Root Mean Square Error	0.116	0.1821

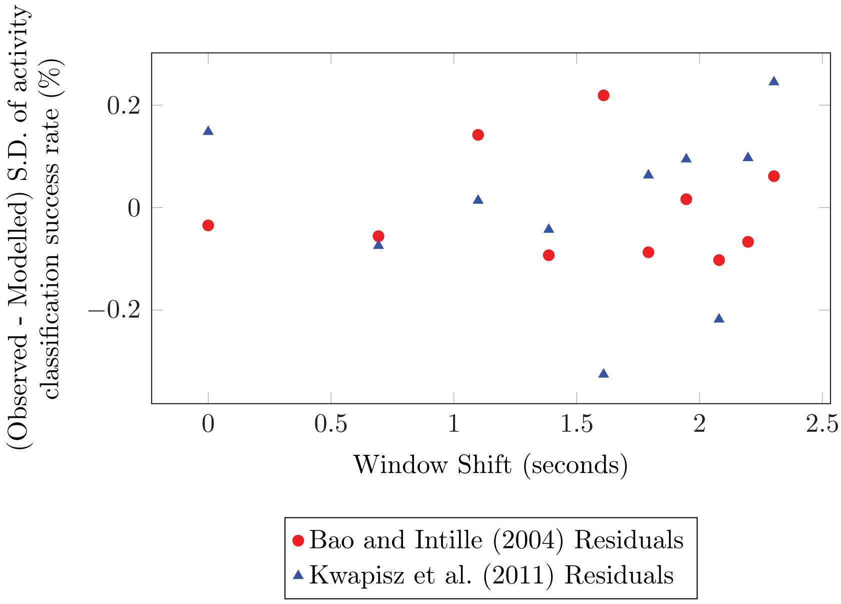


Figure 4.28: Residuals of fitting equation 4.15 and equation 4.16 on the standard deviations of result set success-rates as a function of logarithmic distance between windows for the two feature-sets studied.

From equation 4.15 and equation 4.16, we can say that a linear relationship exists between the standard deviation of activity classification success-rates obtained using 10-fold cross-validation and the logarithm of the distance between windows within the tested range. Hence, an increase in logarithmic window shift leads to an increase in the standard deviation of activity classification success-rates within the tested range.

Discussion

In this section, the impact of window overlaps on activity classification success-rates has been studied. Activity classification has been performed on all gathered activities using two different feature-sets using window overlaps of 0% to 90% in increments of 10% and using a window length of 10 seconds.

The likelihood of window overlaps impacting results from remove-one-subject cross-validation were found to be low. However, the likelihood of impacting results from 10-fold cross-validation was found to be high.

For 10-fold cross-validation, the relationships between mean and standard deviations of success-rates were found to be linear with reference to the logarithm of the distance between windows. However, increases in logarithmic distance between windows lead to a decrease in mean success-rates and an increase in the standard deviation of success-rates.

In other words, increasing window overlap leads to both a higher activity classification success rate, and lower variability in the success rates.

One possible explanation of this observed behaviour is similarities in the statistical properties of signals captured by windows are highest when windows are close and decrease as windows get further apart.

Given a set of windows W from activity A from which feature-vectors in the testing set have been extracted, feature-vectors in the training set that have been extracted from windows that originate from near the W windows, result in an increased similarity between the statistical properties of the training set model of A and the statistical properties of the testing set model of A.

The increased similarity not only results in higher success-rates, but also lower variance in the results. The lower variability is caused by the likelihood of selecting the correct activity when the testing model of the activity and the training model of the activity are similar. The variability increases as the similarity between the training model and the testing model decreases. Since similarity decreases with increase in the distance between windows used to extract feature-vectors in the training model and the testing model.

As an extreme example if the same feature-vectors are used in both the training set and testing testing set we would expect a very high success-rate.

While not extreme, another notable example is when no feature-vectors in the training set come from the same subject as feature-vectors in the testing set. In which case, window overlaps are observed to have little to no impact on success-rates.

It should be noted that the mean success-rate obtained while performing 10-fold cross-validation with no window overlap is higher than the success-rate obtained while performing remove-one-subject cross-validation. In addition, the standard deviation of success-rates obtained while performing 10-fold cross-validation with no window overlap is lower than the standard deviation of success-rates obtained while performing remove-one-subject cross-validation. This implies that while window overlaps result in an increase in success-rates, the presence of the subject's data in both the training set and the testing set also results in increased success-rates.

4.5 Conclusion

In this chapter minimum sampling frequencies, window length and window overlap were studied as to their impact upon activity classification.

First, the MES frequency was defined then extracted based upon success-rates obtained by classifying data downsampled to frequencies in the range of [1Hz,128Hz] in intervals of 1Hz. The minimum sampling requirements of accelerations, rotational velocities and orientations were then compared by comparing the MES frequencies extracted based upon success-rates obtained from each of the three *sources*.

It was found that orientations have the least sampling requirements and rotational velocities have the highest sampling requirements of the three *sources*. However, the differences in sampling requirements, although statistically significant, were found to be small. The largest average difference was observed to be 1.43Hz which was between MES frequencies extracted from classification results of feature vectors extracted from rotational velocities and from classification results of feature vectors extracted from orientations using Kwapisz et al.'s feature-set.

Next, the relationship of length of windows used to extract feature-vectors and the resultant activity classification success-rates was studied. A linear relationship is found between the logarithmic window lengths and the mean resultant activity classification success-rates within the range of window lengths tested. An increase in logarithmic window length results in an increase in the mean activity classification success-rate.

The derivative of the mean success-rates was then studied. The derivative was found to be inversely proportional to the logarithmic window length. The model between mean success-rates and window length was then refined to include this component, and the refined model extrapolated to find the window length at which the models' theoretical prediction would intersect with 100% success-rate. These window lengths were found to be 28.33 seconds for Bao and Intille's feature-set and 31.61 seconds for Kwapisz et al.'s feature-set.

In the discussion section, an explanation for the observed relationship between window lengths and the resultant success-rates was offered. The explanation offered is based on sampling error. Smaller windows have lesser samples hence are impacted with higher sampling errors leading to greater deviations from the population centroid of the captured statistical properties.

Finally, the impact of window overlaps on the resultant mean activity classification success-rates was ascertained. The research found that it is unlikely that window overlaps impact success-rates when remove-one-subject cross-validation is performed. However, it is likely that window overlaps impact mean success-rates when 10-fold cross-validation is performed. The relationship between window overlaps and activity classification success-rates was modelled. The research found that the data fitted a linear relationship between mean success-rates obtained and the logarithmic distance between windows (window shift) for the window shifts studied. An increase in logarithmic window shift results in a decrease in mean success-rates.

The impact of window overlaps on the resultant variance in activity classification success-rates was also studied. Similar to the impact on mean success-rates, the research found that it is unlikely that window overlaps impact the variance in success-rates when remove-one-subject cross-validation is performed. However, it is likely that window overlaps impact the variance in success-rates when 10-fold cross-validation is performed. The relationship between window overlaps and the standard deviation of activity classification success-rates was then modelled. The research found that the data fitted a linear relationship between the standard deviations of success-rates obtained and the logarithmic distance between windows (window shift) for the window shifts studied. An increase in logarithmic window shift results in an increase in the standard deviation of success-rates.

In the discussion section, an explanation for the observed relationship between window overlaps and the mean and standard deviation of the resultant success-rates is offered. The explanation offered is based on the similarity of the training model and the testing model. The closer the sliding windows are to each other, the more similar the data is that goes to creating the training model and the testing model, and hence the more similar the training model is to the testing model. The extreme example of this behaviour are when the same data is used to create both the training and testing models. A notable example is when data from any subject used in the testing set is excluded from the training set. In that case, it is found that window overlaps are unlikely to impact success-rates.

Hence the findings in this chapter can be summarised as follows:

1. Orientations require the lowest sampling frequencies out of the three studied *sources* while rotational velocities require the highest sampling frequencies – the difference however, is small such that it is unlikely to impact data processing or power consumption.
2. There is a depreciating-returns relationship between the length of the windows and activity classification success-rates. This means that longer window lengths always result in higher activity classification success-rates. However, longer window length result in smaller increments in activity classification success-rates.
3. Larger window overlaps result in higher activity classification success-rates

that have lower variance when performing 10-fold cross-validation. However, no impact was observed when performing remove-one-subject cross-validation. The impact observed while using 10-fold cross-validation is likely to be due to similar data existing in both the training set and the testing set (due to the overlaps) and hence increasing the resulting activity classification success-rates.

5 Analysis of the impact of feature-sets, type, number and location of sensors on activity classification accuracy

5.1 Introduction

In this chapter, the relationship between activity classification performance and the different possible choices for feature-sets, monitor configurations used in data gathering, *sources*, and body-locations are studied.

Further analysis of the activity classification performance at a per-activity level including which activities are most easily identifiable and which activities are most confusable with one another, and a comparison of success-rates obtained using 10-fold cross-validation to success-rates obtained using remove-one-subject cross-validation, can be found in chapter 6.

In section 5.3, the activity classification success-rates of the two feature-sets studied, and of the two monitor setups used in data gathering, are studied. This analysis puts into context results obtained in other sections in this chapter and in the rest of the thesis.

In section 5.4, the success-rates obtained when performing activity recognition using accelerations only, rotational velocities only, and orientations only are compared to each other, and to the success-rates obtained when using all three together.

In section 5.5, the success-rates obtained when performing activity recognition using data captured from each body-location are compared. From the analysis, the body-locations are ranked from the body-location that results in the highest success-rates to the body-location that results in the lowest success-rates.

Finally, in section 5.6, success-rates obtained from data captured from different sets of body-locations are compared to each other. The particular combination of body-locations that results in the best performance for a given number of different body-locations are computed. From the success-rates obtained, the relationship of success-rates as a function of number of body-locations monitored is hypothesised.

The chapter is divided into sections. Each section contains research questions dealing with a specific area of interest. Each section begins with a discussion on the importance of studies on the area of interest; research questions in the area of interest are then posed together with reasons why we wish to attempt to answer these particular questions; a methodology of answering the research questions is given; the results of the analysis are then provided and illustrated; and finally, the conclusions and implications of the result findings are discussed. At the end of the chapter, the analysis, findings and implications within findings of the chapter are summarised.

5.2 Overall Methodology

Although the methodology used to answer each question differs slightly, algorithm 6 is a common component used by all methodologies. Algorithm 6 summarises the preprocessing, feature-extraction and classification performed to obtain activity classification results using the feature-set $FeatureSet$, all activities in the set *Activities*, all *sources* in the set *Sources* and all monitors in the set $Monitors$.

Window lengths of 10 seconds were used because 10 seconds was observed to be the maximum window length used in the current activity recognition literature, having only been used by Kwapisz et al. (2011) and Patel et al. (2009). In the analysis of activity recognition literature performed by Lockhart and Weiss (Lockhart & Weiss, 2014), window lengths reported to have been used in activity recognition literature were observed to have a median of 3 seconds and the maximum window length they observed was 10 seconds.

A 50% window overlap was used because it was observed that 50% window overlaps are common within the literature review having been used by Bao and Intille (2004); Figo et al. (2010); He et al. (2008); Krishnan and Panchanathan (2008); Kunze et al. (2005); Preece, Goulermas, Kenney, and Howard (2009); Ravi et al. (2005); Shoaib et al. (2014) and Sun et al. (2010). However, other window overlaps also exist in the literature review including: no overlap ((Kwapisz et al., 2011)), 20% ((Reiss, 2014)), 25% overlap ((Henpraserttae et al., 2011)), 33% overlap ((Lester et al., 2005)).

The next sections provide different parameters and use the results obtained from algorithm 6 to answer the research questions in this chapter.

Algorithm 6 Perform activity classification using the feature-set $FeatureSet$, all activities in the set $Activities$, all sources in the set $Sources$ and all monitors in the set $Monitors$ and return result sets obtained from 10-fold cross-validation.

```
procedure TestConfig(FeatureSet,Activities,Sources,Monitors)
    ResultSet ← []                          ▷ Initialise empty list to hold obtained
                                              confusion matrixes.
    for all Fs ∈ [121Hz, 128Hz] in steps of 1Hz do
        D ← []                              ▷ Initialise feature-vector data set.

        for all A ∈ Activities do
            for all Sub ∈ Subjects who did activity A do
                N ← duration of subject S doing activity A
                for all t ∈ [0, N) in steps of 5s do
                    W ← []                  ▷ Initialise empty feature-vector.
                    for all M ∈ Monitors do
                        for all Src ∈ Sources do
                            Extract feature vector V from data
                                of length L starting from time t
                                from source Src of monitor M
                                on subject Sub doing activity A
                                downsampled to frequency Fs
                                using feature-set FeatureSet
                        end for
                        Append V to W
                    end for
                    Add W to D
                end for
            end for
        end for

        F ← Split D into 10 folds
        for all f ∈ F do
            TrainingSet ← all folds except for f
            Train J48 classifier with TrainingSet
            Test with f
            Add success-rate to ResultSet
        end for
    end for

    return ResultSet
end procedure
```

5.3 Performance comparison of feature-sets and the 3 monitor vs the 6 monitor setup

In this section, we are interested in comparing the activity classification success-rates obtained from feature-vectors computed using the two feature-sets, so as to learn how the two feature-sets compare with each other.

The research question posed in this section is: how do activity classification success-rates obtained using Bao and Intille's feature-set compare to those obtained using Kwapisz et al.'s feature-set?

Comparing the performance of the two feature-sets is important because it puts into context results obtained in the rest of the chapter when comparing the performance of different subsets of the gathered data using the two feature-sets.

In addition, because of how the data was gathered, two setups are possible:

1. Using all activities but including only 3 monitors: thigh (or phone), chest and dominant wrist (referred to as "the 3 monitor setup").
2. Using all monitors but excluding walking and running activities (referred to as "the 6 monitor setup").

Therefore, an additional dimension studied in this section is the 3 monitor setup compared to the 6 monitor setup. Hence, an additional research question posed in this section is: how do success-rates obtained using the 3 monitor setup compare to those obtained using the 6 monitor setup?

Methodology

The data gathered includes activities with six monitors mounted on each subject and also activities with 3 monitors mounted on each subject (see section 3.3 for more details on the activities gathered using six monitors and activities gathered using three monitors). The activities with only 3 monitors mounted are walking and running only. The remaining 19 (nineteen) activities were gathered using six monitors. The three monitors were mounted on the thigh, chest and dominant wrist while gathering data for the activities walking and running. These three locations are a subset of the six monitors mounted on all other activities. Table 5.1 summarises the data available for each activity with reference to where the monitors were mounted on the subjects.

Table 5.1: Body-locations on which monitors were mounted for each activity gathered.

Activity	***Body-location data available***					
	A	**C**	**DUA**	**DW**	**NDW**	**T**
Walking (Flat Surface)	✗	✓	✗	✓	✗	✓
Running	✗	✓	✗	✓	✗	✓
Brushing Teeth	✓	✓	✓	✓	✓	✓
Dicing	✓	✓	✓	✓	✓	✓
Dusting	✓	✓	✓	✓	✓	✓
Folding Clothes	✓	✓	✓	✓	✓	✓
Grating	✓	✓	✓	✓	✓	✓
Ironing	✓	✓	✓	✓	✓	✓
Peeling Veg.	✓	✓	✓	✓	✓	✓
Stiring	✓	✓	✓	✓	✓	✓
Sweeping	✓	✓	✓	✓	✓	✓
Talking (Phone)	✓	✓	✓	✓	✓	✓
Texting (Phone)	✓	✓	✓	✓	✓	✓
Using PC	✓	✓	✓	✓	✓	✓
Vacuuming	✓	✓	✓	✓	✓	✓
Walking D. Stairs	✓	✓	✓	✓	✓	✓
Walking U. Stairs	✓	✓	✓	✓	✓	✓
Washing Dishes	✓	✓	✓	✓	✓	✓
Washing Hands	✓	✓	✓	✓	✓	✓
Washing Veg.	✓	✓	✓	✓	✓	✓
Watching TV	✓	✓	✓	✓	✓	✓
Writing	✓	✓	✓	✓	✓	✓

Hence, comparison is performed for the two feature-sets using six monitors but not including walking and running (6 monitor setup), and for three monitors and including walking and running (3 monitor setup).

To that end, four sets of results were computed using algorithm 6. The parameters provided to algorithm 6 for the result-sets are given in table 5.2.

Results

The result sets obtained using algorithm 6 and the parameters given in table 5.2 are summarised as figure 5.1.

Figure 5.1 shows normalised histograms of the success-rates obtained when activity classification is performed using algorithm 6 using parameters given in table 5.2. The success-rates obtained from algorithm 6 include success-rates of each fold and of sampling frequencies in the range [121Hz,128Hz].

The result sets are observed to fit a Gaussian model with mean success-rates 76.4%, 75.1%, 79.0% and 78.7% and standard deviation 1.59%, 1.62%, 1.68% and 1.62% for the result sets *Bao-3*, *Kwapisz-3*, *Bao-6* and *Kwapisz-6* respectively. The Gaussian models were fitted using maximum-likelihood estimation .

Table 5.3 gives results obtained by comparing the result sets. The values (in the column order) are:

1. The t-statistic from a paired two-sample two-tailed t-test with $\alpha = 0.05$ between the two result sets being compared testing the null hypothesis that the two result sets came from independent random samples from normal distributions with equal means and unknown variances. The test did not assume that the two result sets had equal variance by using Satterthwaite's approximation of the effective degrees of freedom. All the tests rejected the null hypothesis. Therefore, the result sets compared had a statistically significant difference in mean.
2. The F-statistic from a two-sample two-tailed F-test with $\alpha = 0.05$ between

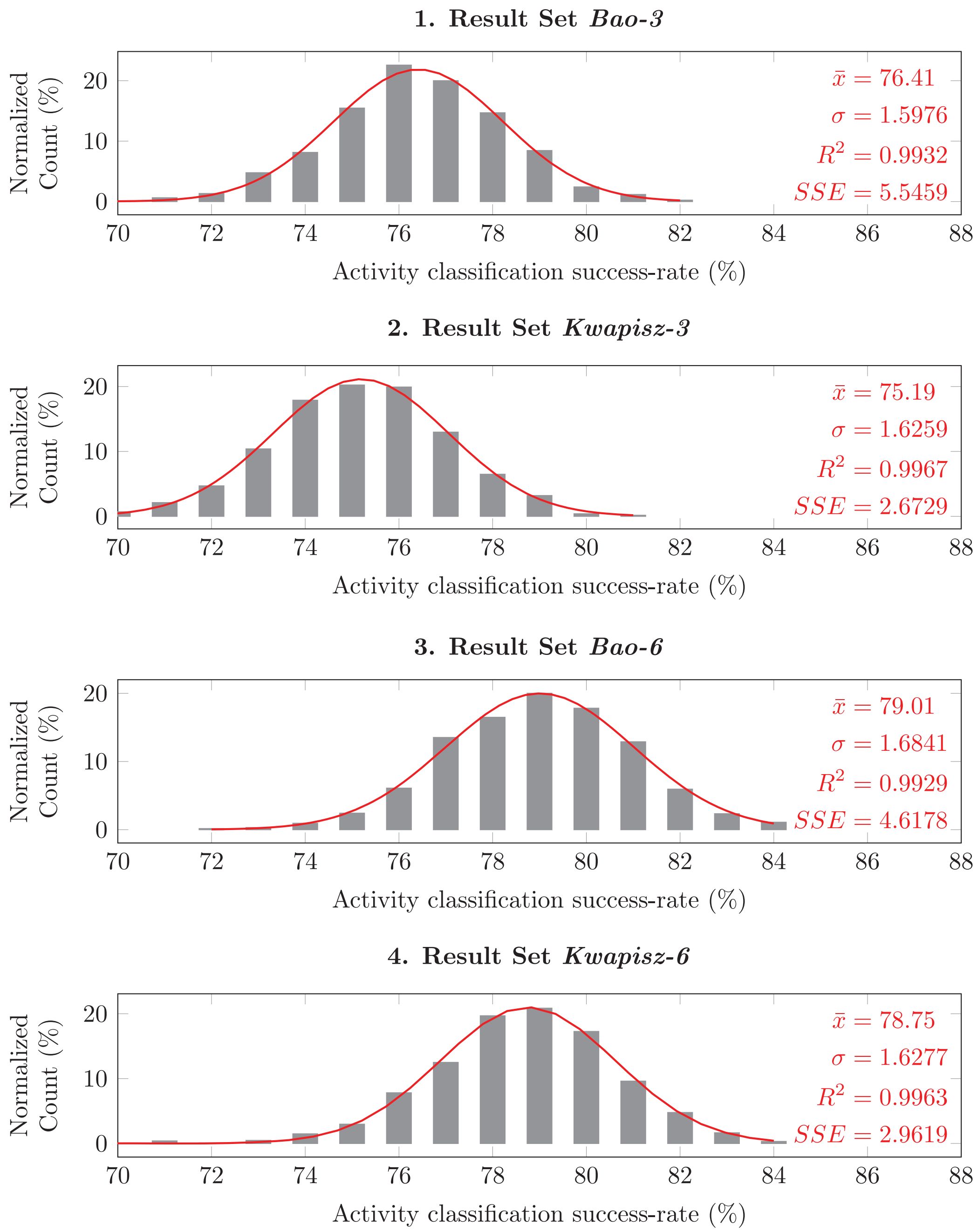


Figure 5.1: Normalised histograms of activity classification success-rates obtained using all three *sources* (accelerations, rotational velocities and orientations). Histograms 1 and 2 are of success-rates obtained using the 3 monitor setup. Histograms 3 and 4 are of success-rates obtained using the 6 monitor setup. Success-rates plotted in histogram 1 and 3 are obtained using Bao and Intille's feature-set while those in histogram 2 and 4 are obtained using Kwapisz et al.'s feature-set.

the two result sets being compared testing the null hypothesis that the two result sets came from independent random samples from normal distributions with equal variances. All the tests rejected the null hypothesis. Therefore, the result sets compared have a statistically significant difference in variance.

3. The difference in the standard deviations of the result sets tested. Although the F-test found that the differences in the variances of the result sets tested are statistically significant, the differences in standard deviation are seen to be low, in all cases being less than 1%.
4. The population mean of differences of the two result sets, obtained by computing the difference of the two result sets compared for each sample then fitting a Gaussian model and extracting the mean of the model. The differences of success-rates and fitted Gaussian models are shown as figure 5.2, figure 5.3, figure 5.4 and figure 5.5.
5. The population standard deviation of differences of the two result sets, obtained by computing the difference of the two result sets compared for each sample then fitting a Gaussian model and extracting the standard deviation of the model.
6. The probability of a sample from the first result set having a higher success-rate than the equivalent sample in the second result when when using a paired test. The probability is computed as the integral of the PDF of the fitted Gaussian model that is below zero.

Differences in mean

The mean differences are found to be less than 4% for all result sets compared. In addition, the mean differences are found to be lower when comparing result sets obtained from different feature-sets but similar monitor setup (i.e. *Bao-3* vs *Kwapisz-3* and *Bao-6* vs *Kwapisz-6*), and particularly low when comparing with the 6 monitor setup (i.e. *Bao-6* vs *Kwapisz-6*). For both monitor setups, Bao

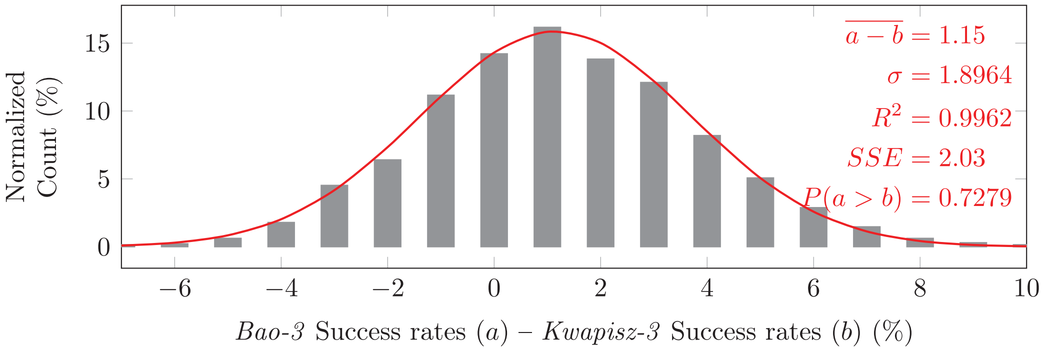


Figure 5.2: Normalised histogram of the differences in activity classification success-rates in result set *Bao-3* from those of *Kwapisz-3*.

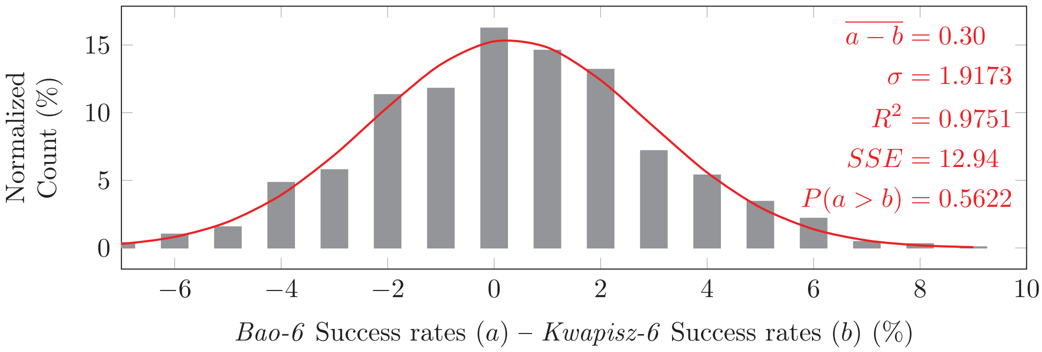


Figure 5.3: Normalised histogram of the differences in activity classification success-rates in result set *Bao-6* from those of *Kwapisz-6*.

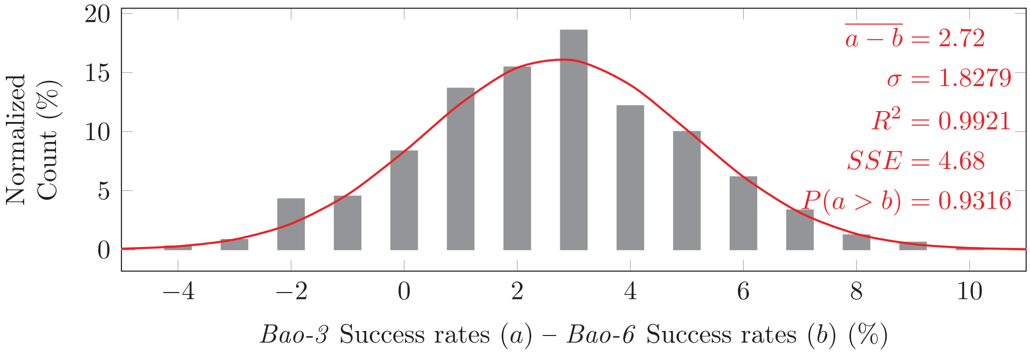


Figure 5.4: Normalised histogram of the differences in activity classification success-rates in result set *Bao-6* from those of *Bao-3*.

and Intille's feature-set resulted in slightly higher success-rates. From the fitted Gaussian models, the probability of having a higher success-rate using Bao and Intille's feature-set than than using Kwapisz et al.'s feature-set is 0.72 and 0.56 using the 3 monitor setup and 6 monitor setup respectively. Notice that 0.56 is

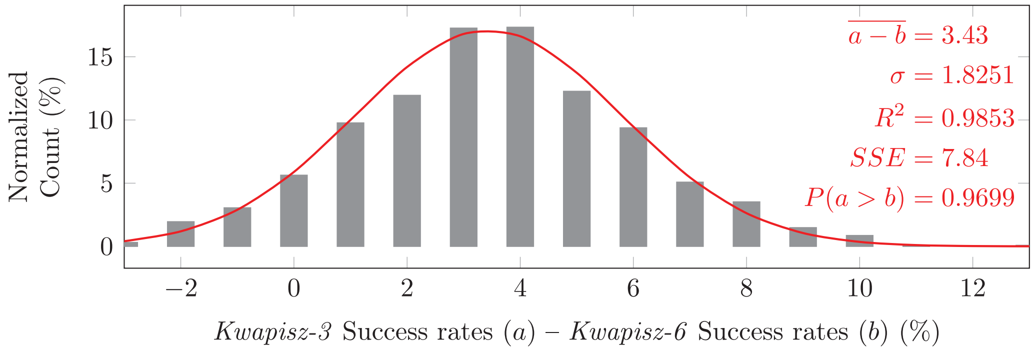


Figure 5.5: Normalised histogram of the differences in activity classification success-rates in result set *Kwapisz-6* from those of *Kwapisz-3*.

closer to chance (0.5).

The mean differences where the feature-set is similar but the monitor setup is different are found to be higher (i.e. *Bao-3* vs *Bao-6* and *Kwapisz-3* vs *Kwapisz-6*), with Kwapisz et al.'s feature-set resulting in a higher mean difference than Bao and Intille's feature-set. For both feature-sets, the 6 monitor setup resulted in higher success-rates than the 3 monitor setup. From the fitted Gaussian models, the probability of having a higher success-rate using the 6 monitor setup than using the 3 monitor setup is 0.9316 and 0.9699 while using Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively.

Differences in standard deviation

The differences in standard deviation between the feature-sets were found to be small but statistically significant. The difference in standard deviation between 3 and 6 monitor setup result sets obtained using Bao and Intille's feature-set are found to be larger than those obtained using Kwapisz et al.'s feature-set. The difference in standard deviations between the two feature-sets was found to be lower while using the 3 monitor setup than while using the 6 monitor setup.

Discussion

In this section, the success-rates obtained by using Bao and Intille's feature-set have been compared to those obtained by using Kwapisz et al.'s feature-set.

Unfortunately, the walking (on a flat surface) and running activities in the data gathered includes only three monitors mounted on the subjects while the rest of the activities include six monitors mounted on the subjects.

Due to this difference, an additional comparison dimension was added. An additional comparison was performed between the two setups: one that excludes walking and running but includes all six monitors available, and other that includes all activities but excludes other monitors except for the three monitors available for walking and running (thigh or phone, chest and dominant wrist).

Success-rates obtained using Bao and Intille's feature-set were found to be higher than those obtained using Kwapisz et al.'s feature-set. However, the differences are observed to average 1.15% and 0.30% only for the 3 monitor setup and 6 monitor setup respectively.

Based on the fitted Gaussian models, the probability of attaining a higher success-rate while using Bao and Intille's feature-set than while using Kwapisz et al.'s feature-set is 0.7279 while using the 3 monitor setup. The equivalent probability while using the 6 monitor setup is 0.5633 and is closer to 0.50 (chance).

Success-rates of Bao and Intille's feature-set were found to vary more than those of Kwapisz et al.'s feature-set while using the 6 monitor setup, but were found to vary less than those of Kwapisz et al's feature-set while using the 3 monitor setup. The difference in standard deviations, however, was observed to be less than 1% for both feature-sets and hence considered low.

Comparing the 3 monitor setup to the 6 monitor setup, it was found that success-rates obtained from the 6 monitor setup are higher and vary more than those obtained from the 3 monitor setup when both Bao and Intille's feature-set and Kwapisz et al.'s feature-set were used. The difference in standard deviation between the two sets are less than 1% for both feature-sets studied and hence considered low. The differences in the means of the result sets was found to be 2.72% and 3.43% while using Bao and Intille's and Kwapisz et al.'s feature-set respectively.

Based on the fitted Gaussian models, the probability of attaining a higher success-rate using the 6 monitor setup than while using the 3 monitor setup is 0.9316 and 0.9699 for Bao and Intille's and Kwapisz et al.'s feature-set respectively.

Hence, from the results, we can summarise that the examined differences in monitor setup had a higher impact on obtained results with the 6 monitor setup resulting into higher success-rates than those obtained by the 3 monitor setup. The difference is less than 4% for either feature-set used but is more prevalent, with a much higher chance of obtaining a higher success-rate while using the 6 monitor setup than while using the 3 monitor setup.

The impact of the feature-sets on the obtained results is observed to be less resulting in smaller differences and in smaller probabilities of there being a difference. Bao and Intille's feature-set results in slightly better success-rates than Kwapisz et al.'s feature-set. The difference is lesser when the 6 monitor setup is used than when the 3 monitor setup is used.

A possible explanation of why the 6 monitor setup results in higher success-rates is that when 6 monitors are used, more information is available to distinguish activities from each other. However, since the two setups have different activities, the difference could also be due to walking and running activities, which are included in the 3 monitor setup but not in the 6 monitor setup. It could also be a combination of both factors contributing to the difference between the two monitor setups. In this section, we can not deduce the contribution of each of the two factors since the comparison included both different numbers of activities in the two monitor setups and different numbers of monitors in the two monitor setups.

A deeper analysis of the impact of number os monitors on activity classification success-rates is performed in section 5.6. The analysis performed in section 5.6 will analyse the obtained success-rates as the number of monitors mounted on different locations on the subject increases and hence provide a better understanding of the impact of the number of different body-locations used on activity classification success-rates.

A deeper analysis of the activity classification accuracy of each activity is performed in section 6.2. The analysis performed in section 6.2 will compare the accuracies of different activities in order to understand which activities are more easily identified than other activities.

Table 5.2: Parameters provided to algorithm 6 so as to compare the activity classification success-rates obtained while using the two feature-sets studied.

Result Set	***FeatureSet***	***Activities***	***Sources***			***Monitors***					
			Accel.	**Rot. Vel.**	**Orient.**	**A**	**C**	**DUA**	**DW**	**NDW**	**T**
Bao-3	Bao and Intille (2004)	All	✓	✓	✓		✓		✓		✓
Kwapisz-3	Kwapisz et al. (2011)	All	✓	✓	✓		✓		✓		✓
Bao-6	Bao and Intille (2004)	All except run & walk	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kwapisz-6	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 5.3: Statistics obtained while comparing result sets.

Result Sets Compared (a vs b)	t-**statistic**	F-**statistic**	$\boldsymbol{\sigma_a - \sigma_b}$ (%)	$\overline{\boldsymbol{a-b}}$ (%)	$\boldsymbol{\sigma_{a-b}}$ (%)	$\boldsymbol{P(a>b)}$
Bao-3 vs *Kwapisz-3*	18.2537	0.8250	−0.0283	1.15	1.8964	0.7279
Bao-6 vs *Kwapisz-6*	5.0448	0.974	0.0564	0.30	1.9173	0.5622
Bao-6 vs *Bao-3*	−48.5181	0.9157	0.0865	2.72	1.8279	0.9316
Kwapisz-6 vs *Kwapisz-3*	−58.8013	1.0811	0.0018	3.43	1.8251	0.9699

5.4 Performance comparison of accelerations, rotational velocities and orientations

In this section we are interested in comparing the performance of each *source* to each other *source*, and to the performance of the three *sources* when combined.

The research questions asked are:

1. How do the success-rates obtained from feature-vectors extracted from each *source* compare to success-rates obtained from feature-vectors extracted from other *sources*?
2. How do the success-rates obtained from feature-vectors extracted from each *source* compare to success-rates obtained from feature-vectors extracted from all three *sources* combined?

Knowing the relative performance of each *source* in comparison to other *sources* and combination of *sources* would allow activity recognition researchers a better understanding while selecting which *source* to use for activity recognition.

Each *source* requires the use of a different sensor or a set of sensors: accelerations require an accelerometer; rotational velocities require a gyroscope; and orientations require an accelerometer, gyroscope and magnetometer (compass). Hence, the selection of the *source* to use for activity classification has implications on the sensing node's hardware requirements.

Methodology

The data gathered includes activities with 6 monitors mounted on each subject and also activities with 3 monitors mounted on each subject. The activities with only 3 monitors mounted are walking and running only. The rest of the activities were gathered using 6 monitors. Section 5.3 found that the monitor setups have a significant impact on results obtained.

Therefore, to answer the research questions, result sets were computed for both setups: all 6 monitors but excluding walking and running (6 monitor setup) and all activities but including only the monitors mounted on the thigh (or phone), chest and dominant wrist (3 monitor setup). Refer to section 5.3 for a comparison of the impact of the 6 monitor setup and the 3 monitor setup on activity classification success-rates.

12 sets of results were computed using algorithm 6. The parameters provided to algorithm 6 for the result sets are given in table 5.4.

In addition, the result sets from section 5.3, where all the *sources* were combined were used to compare the results obtained from each *source* to the results obtained with the combined *sources*. As in section 5.3, the result sets are refered to as *Bao-3*, *Bao-6*, *Kwapisz-3* and *Kwapisz-6*. Refer to section 5.3 for further details about these result sets.

Results

The result sets obtained using algorithm 6 and the parameters given in table 5.4 are summarised as figure 5.6 for the 6 monitor setup result sets and figure 5.7 for the 3 monitor setup result sets.

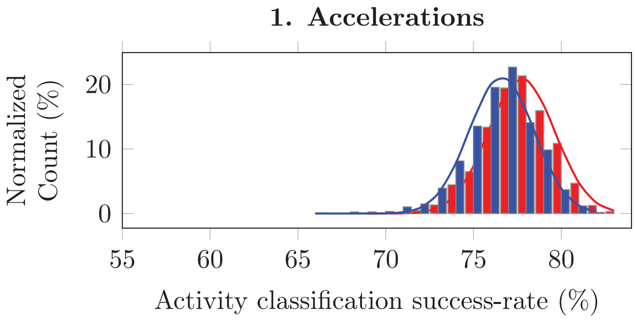
Figure 5.6 and figure 5.7 show normalised histograms of the success-rates obtained when activity classification is performed using algorithm 6 using parameters given in table 5.4.

The result sets are observed to fit Gaussian models. The means and standard deviations of the best fitting Gaussian models are given in table 5.5.

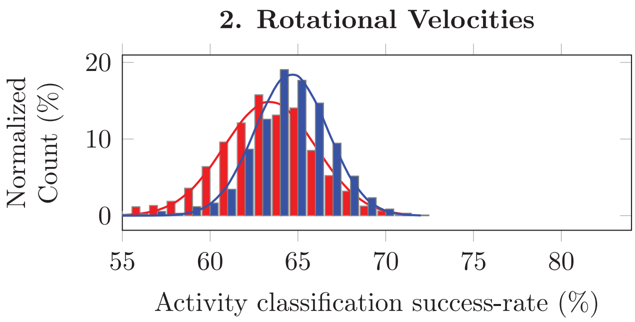
From the data given in table 5.5, we can notice that the mean success-rates of result sets obtained from accelerations are higher than those obtained from either rotational velocities or orientations for the same monitor setup and feature-set. In addition, mean success-rates of result sets obtained from orientations are higher than

Table 5.4: Parameters provided to algorithm 6 so as to compare the activity classification success-rates obtained from each *source*.

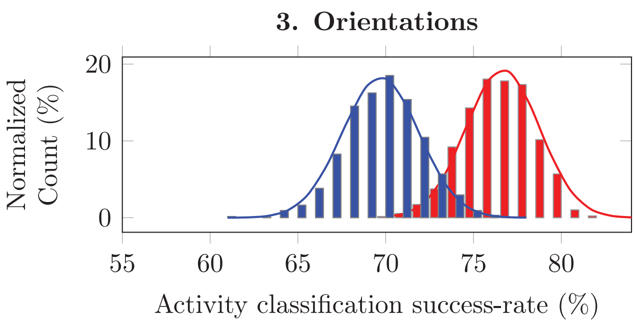
Result Set	*FeatureSet*	*Activities*	*Sources*			*Monitors*					
			Accel.	Rot. Vel.	Orient.	A	C	DUA	DW	NDW	T
Bao-Accel-3	Bao and Intille (2004)	All	✓				✓		✓		✓
Bao-Gyro-3	Bao and Intille (2004)	All		✓			✓		✓		✓
Bao-Orient-3	Bao and Intille (2004)	All			✓		✓		✓		✓
Bao-3	Bao and Intille (2004)	All	✓	✓	✓		✓		✓		✓
Bao-Accel-6	Bao and Intille (2004)	All except run & walk	✓			✓	✓	✓	✓	✓	✓
Bao-Gyro-6	Bao and Intille (2004)	All except run & walk		✓		✓	✓	✓	✓	✓	✓
Bao-Orient-6	Bao and Intille (2004)	All except run & walk			✓	✓	✓	✓	✓	✓	✓
Bao-6	Bao and Intille (2004)	All except run & walk	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kwapisz-Accel-3	Kwapisz et al. (2011)	All	✓				✓		✓		✓
Kwapisz-Gyro-3	Kwapisz et al. (2011)	All		✓			✓		✓		✓
Kwapisz-Orient-3	Kwapisz et al. (2011)	All			✓		✓		✓		✓
Kwapisz-3	Kwapisz et al. (2011)	All	✓	✓	✓		✓		✓		✓
Kwapisz-Accel-6	Kwapisz et al. (2011)	All except run & walk	✓			✓	✓	✓	✓	✓	✓
Kwapisz-Gyro-6	Kwapisz et al. (2011)	All except run & walk		✓		✓	✓	✓	✓	✓	✓
Kwapisz-Orient-6	Kwapisz et al. (2011)	All except run & walk			✓	✓	✓	✓	✓	✓	✓
Kwapisz-6	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓	✓	✓	✓	✓	✓	✓



	Bao-Accel-6	***Kwapisz-Accel-6***
$\bar{x}$	77.74	76.60
σ	1.6381	1.6162
R^2	0.9930	0.9870
SSE	4.89	10.93



	Bao-Gyro-6	***Kwapisz-Gyro-6***
$\bar{x}$	63.40	64.62
σ	1.9183	1.7112
R^2	0.9847	0.9939
SSE	9.12	5.33



	Bao-Orient-6	***Kwapisz-Orient-6***
$\bar{x}$	76.66	69.71
σ	1.7236	1.7565
R^2	0.9891	0.9951
SSE	6.7534	3.0780

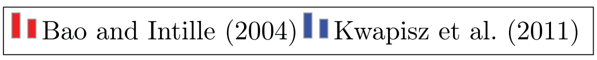
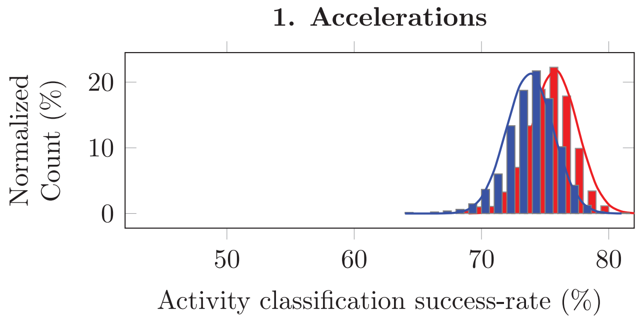
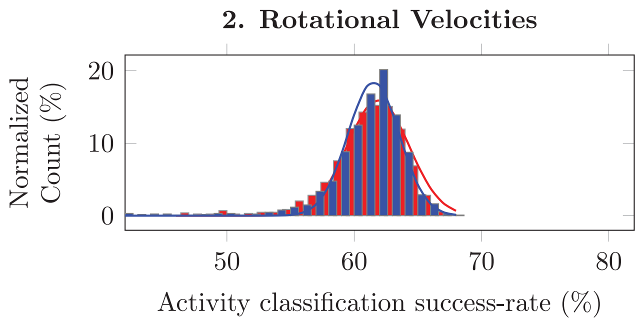


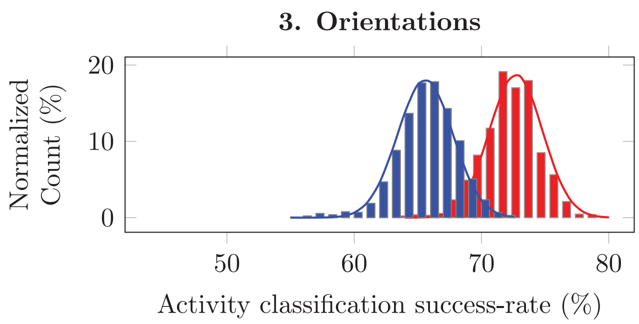
Figure 5.6: Normalised histograms of activity classification success-rates obtained using each individual *source*: accelerations (upper), rotational velocities (middle) and orientations (lower). Each histogram displays results obtained using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set for the 6-monitor setup. Classification was performed using all monitors but excluded walking and running activities.



	Bao-Accel-3	*Kwapisz-Accel-3*
$\bar{x}$	75.76	73.79
σ	1.5947	1.6083
R^2	0.9940	0.9954
SSE	4.81	3.92



	Bao-Gyro-3	*Kwapisz-Gyro-3*
$\bar{x}$	61.87	61.52
σ	1.8551	1.6957
R^2	0.9867	0.9800
SSE	8.79	17.36



	Bao-Orient-3	*Kwapisz-Orient-3*
$\bar{x}$	72.70	65.58
σ	1.7203	1.7515
R^2	0.9743	0.9982
SSE	18.84	1.22

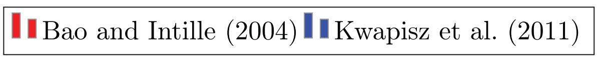


Figure 5.7: Normalised histograms of activity classification success-rates obtained using each individual *source*: accelerations (upper), rotational velocities (middle) and orientations (lower). Each histogram displays results obtained using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set for the 3-monitor setup. Classification was performed using all activities but using only 3 monitors: thigh (or phone), chest and dominant wrist.

Table 5.5: Mean and standard deviations of Gaussian models fitted onto the result sets obtained from parameters given in table 5.4

Result Set	$\bar{x}$	σ
Bao-Accel-6	77.7	1.63
Bao-Orient-6	76.6	1.72
Kwapisz-Accel-6	76.6	1.61
Bao-Accel-3	75.7	1.59
Kwapisz-Accel-3	73.7	1.60
Bao-Orient-3	72.7	1.72
Kwapisz-Orient-6	69.7	1.75
Kwapisz-Orient-3	65.5	1.75
Kwapisz-Gyro-6	64.6	1.71
Bao-Gyro-6	63.4	1.91
Bao-Gyro-3	61.8	1.85
Kwapisz-Gyro-3	61.5	1.69

those obtained from rotational velocities for the same monitor setup and feature-set. As observable in figure 5.6 and figure 5.7, overlaps exist between result sets obtained from different *sources* but of the same monitor setup and feature-set.

A paired two-sample two-tailed t-test with $\alpha = 0.05$ between all possible pairings of the result sets was run to test the null hypothesis that the pairs of result sets came from independent random samples from normal distributions with equal means and unknown variances. The test did not assume that the two result sets had equal variance by using Satterthwaite's approximation of the effective degrees of freedom.

The null hypothesis was rejected for all but one pair of result sets: *Kwapisz-Accel-6* and *Bao-3* (results obtained using Bao and Intille's feature-set from feature-sets derived from accelerations, rotational velocities and orientations). The results of the p-values obtained from the tests are summarised in the box-plot figure 5.8.

The failure to reject the null hypothesis between the result sets *Kwapisz-Accel-6* and *Bao-3* means that there is not enough evidence to support the claim that the

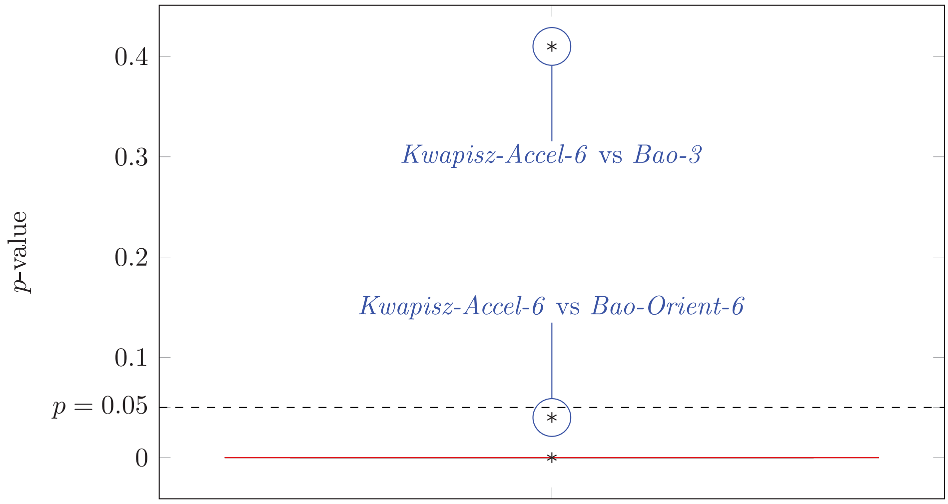


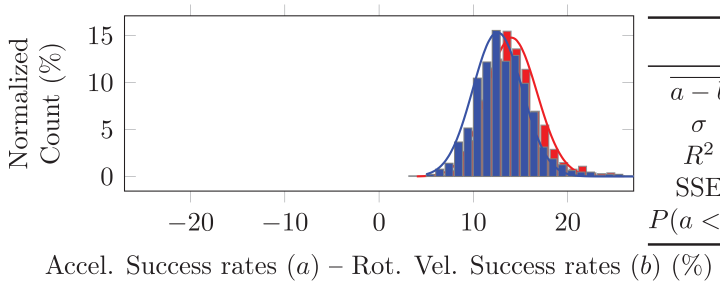
Figure 5.8: Box plot of p-values obtained from paired two-tailed two-sample t-tests between all possible pairs of result sets obtained by calling algorithm 6 using the parameters given in table 5.4. The paired two-tailed two-sample t-tests tested the null hypothesis that pairs of result sets came from independent random samples from normal distributions with equal means and unknown variances. At the preselected significance level of 0.05, a p-value of 0.05 or below results in the rejection of the null hypothesis, while a p-value above 0.05 results in the failure to reject the null hypothesis. Most of the p-values are very close to zero, hence the line at the bottom of the figure represents most of the p values obtained. Outliers are shown as asterisks. *Bao-3* refers to results obtained using Bao and Intille's feature-set from feature-sets derived from accelerations, rotational velocities and orientations

two samples differ.

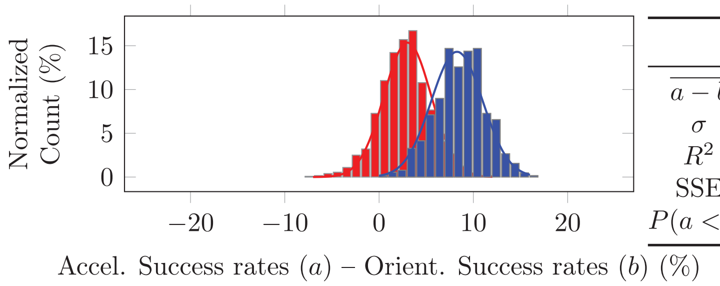
Accelerations vs Rotational Velocities vs Orientations for the 3 monitor setup

The differences of the success-rates obtained from each *source* from those obtained from each other *source* for the 3 monitor setup are summarised in figure 5.9.

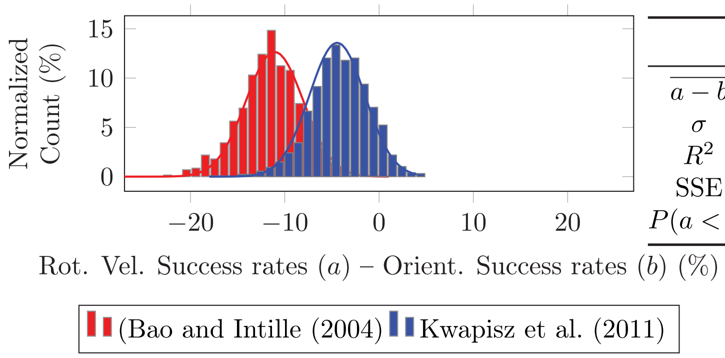
Figure 5.9 shows the distribution of the differences between the success-rates obtained from feature-vectors extracted from rotational velocities from those of feature-vectors extracted from accelerations, success-rates obtained from feature-vectors extracted from orientations from those of feature-vectors extracted from accelerations, and success-rates obtained from feature-vectors extracted from orientations from those of feature-vectors extracted from rotational velocities, for the 3



	Bao	***Kwapisz***
$\overline{a-b}$	14.05	12.56
σ	1.9488	1.9254
R^2	0.9878	0.9891
SSE	7.14	6.32
$P(a<b)$	~ 0.0	~ 0.0



	Bao	***Kwapisz***
$\overline{a-b}$	3.01	8.26
σ	1.8828	1.9655
R^2	0.9941	0.9575
SSE	3.63	20.10
$P(a<b)$	0.0549	~ 0.0



	Bao	***Kwapisz***
$\overline{a-b}$	-11.07	-4.44
σ	2.0778	2.0426
R^2	0.9886	0.9945
SSE	5.32	2.59
$P(a<b)$	~ 1.0	0.9851

Figure 5.9: Histograms of differences in success-rates of each *source* from each those of each other *source* of the 3 monitor setup. Differences in accelerometer success-rates and rotational velocities (upper), accelerometer success-rates and orientation success-rates (middle) and rotational velocities and orientations (lower) are shown.

monitor setup and for both feature-sets studied.

The average differences between success-rates obtained from accelerations and those obtained from rotational velocities are 14.05% and 12.56% for Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively. Those between success-rates obtained from accelerations and those obtained from orientations are 3.01% and 8.26%, and between success-rates obtained from rotational velocities and those obtained from orientations are 11.07% and 4.44%.

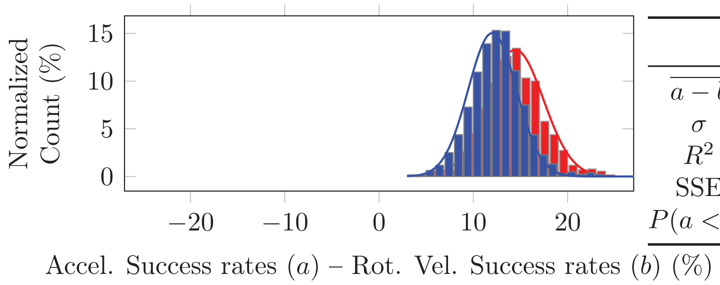
From the fitted Gaussian models, the probability of accelerations resulting in lower success-rates than rotational velocities is observed to be very low for both feature-sets studied. The probability of accelerations resulting in lower success-rates than orientations is also low, although it is 5.59% for Bao and Intille's feature-set. The probability of rotational velocities resulting in lower success-rates than orientations is observed to be high and almost certain for Bao and Intille's feature-set.

Implications of the results of this analysis will be discussed in the discussion section (section 5.4).

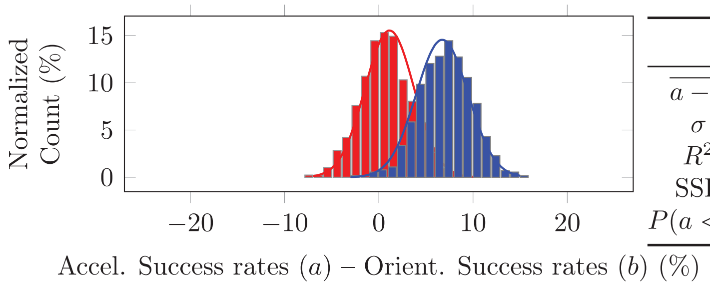
Accelerations vs Rotational Velocities vs Orientations for the 6 monitor setup

The differences of the success-rates obtained from each *source* from those obtained from each other *source* for the 6 monitor setup are summarised in figure 5.10.

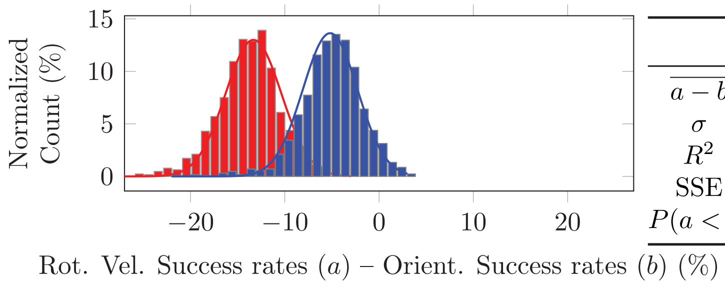
Figure 5.10 shows the distribution of the differences between the success-rates obtained from feature-vectors extracted from rotational velocities from those of feature-vectors extracted from accelerations, success-rates obtained from feature-vectors extracted from orientations from those of feature-vectors extracted from accelerations, and success-rates obtained from feature-vectors extracted orientations from those of feature-vectors extracted from rotational velocities, for the 6 monitor setup and for both feature-sets studied.



	Bao	***Kwapisz***
$\overline{a-b}$	14.44	12.01
σ	2.0565	1.9030
R^2	0.9888	0.9978
SSE	5.24	1.38
$P(a<b)$	~ 0.0	~ 0.0



	Bao	***Kwapisz***
$\overline{a-b}$	1.13	6.72
σ	1.9199	1.9738
R^2	0.9952	0.9953
SSE	2.48	2.15
$P(a<b)$	0.2781	~ 0.0



	Bao	***Kwapisz***
$\overline{a-b}$	-13.32	-5.18
σ	2.0460	2.0253
R^2	0.9884	0.9942
SSE	5.78	3.03
$P(a<b)$	~ 1.0	0.9947

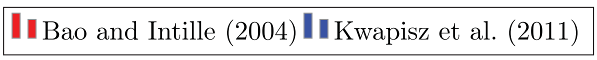


Figure 5.10: Histograms of differences in success-rates of each *source* from each those of each other *source* of the 6 monitor setup. Differences in accelerometer success-rates and rotational velocities (upper), accelerometer success-rates and orientation success-rates (middle) and rotational velocities and orientations (lower) are shown.

The average differences between success-rates obtained from accelerations and those obtained from rotational velocities are 14.44% and 12.01% for Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively. Those between success-rates obtained from accelerations and those obtained from orientations are 1.13% and 6.72%, and between success-rates obtained from rotational velocities and those obtained from orientations are 13.32% and 5.18%. It should be noted that these average differences are similar to those obtained for the 3 monitor setup.

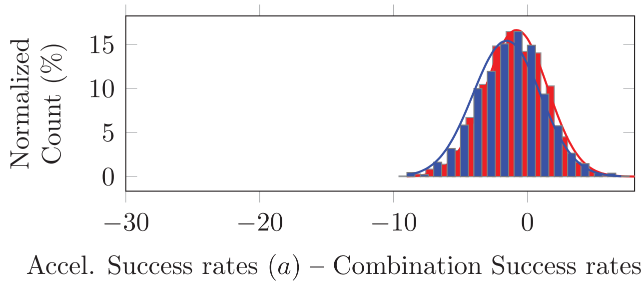
From the fitted Gaussian models, the probability of accelerations resulting in lower success-rates than rotational velocities is observed to be very low for both feature-sets studied. The probability of accelerations resulting in lower success-rates than orientations is also low for Kwapisz et al.'s feature-set but is observed to be 0.2781 for Bao and Intille's feature-set. The probability of rotational velocities resulting in lower success-rates than orientations is observed to be high and almost certain for Bao and Intille's feature-set. It should be noted that with the exception of the increased probability of acceleration success-rates being lower than orientations, the rest of the probabilities are very similar to those obtained for the 3 monitor setup.

Implications of the results of this analysis will be discussed in the discussion section (section 5.4).

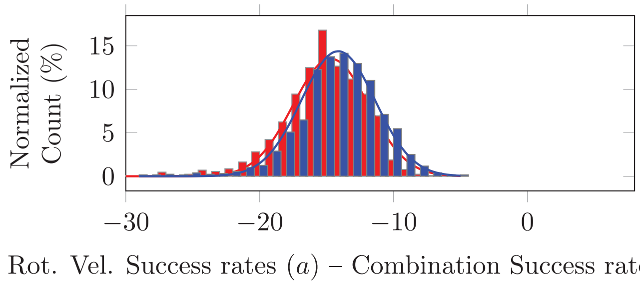
Each *source* vs the combination of the three *sources* for the 3 monitor setup

Success-rates obtained using feature-vectors extracted the combination of the three *sources* for the 3 monitor setup are shown in figure 5.1. Success-rates obtained using feature-vectors extracted from each of the three *sources* for the 3 monitor setup are shown in figure 5.7.

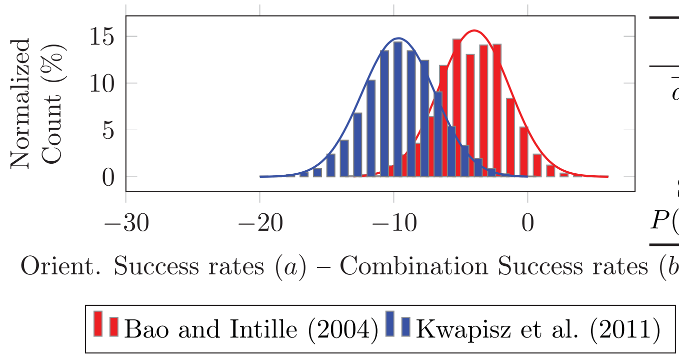
The differences of the success-rates obtained from each *source* from those obtained from the combination of the sources *source* (*Bao-3* and *Kwapisz-3*) for the



	Bao	***Kwapisz***
$\overline{a-b}$	-0.81	-1.55
σ	1.8125	1.8872
R^2	0.9878	0.9949
SSE	3.32	2.82
$P(a<b)$	0.6725	0.7943



	Bao	***Kwapisz***
$\overline{a-b}$	-14.68	-14.13
σ	1.9966	1.9756
R^2	0.98	0.9942
SSE	10.90	3.43
$P(a<b)$	~ 1.0	~ 1.0



	Bao	***Kwapisz***
$\overline{a-b}$	-3.99	-9.65
σ	1.9352	1.9614
R^2	0.9918	0.9980
SSE	4.81	1.07
$P(a<b)$	0.9804	~ 1.0

Figure 5.11: Histograms of differences in success-rates obtained from feature-vectors extracted from each *source* from those obtained from feature-vectors extracted from all three *sources* of the 3 monitor setup. Differences in accelerometer success-rates (upper), rotational velocities (middle) and orientations (lower) are shown.

3 monitor setup are summarised in figure 5.11.

Figure 5.11 shows the distribution of the differences between the success-rates obtained from feature-vectors extracted from a combination of all three *sources* from those of feature-vectors extracted from accelerations only, rotational velocities only and orientations only for the 3 monitor setup and for both feature-sets studied.

On average for the 3 monitor setup, feature-vectors from accelerations only result in success-rates that are 0.81% and 1.55% lower than feature-vectors from all three *sources* combined for Bao and Intille's feature-set and Kwapisz et al.'s feature-sets. Feature-vectors from rotational velocities result in success-rates that are 14.68% and 14.13% lower than feature-vectors from all three *sources* combined. Feature-vectors from orientations result in success-rates that are 3.99% and 9.65% lower than feature-vectors from all three *sources* combined.

From the fitted Gaussian models, the probability of feature-vectors extracted from rotational velocities and those from orientations resulting in lower success-rates than feature-vectors extracted from the three *sources* combined, is observed to be high to almost certain for both feature-sets. However, the probability of feature-vectors extracted from accelerations resulting in lower success-rates than feature-vectors extracted from the three *sources* combined is lower than any of the other *sources* but above chance with Bao and Intille's feature-set having a probability of 0.6725 and Kwapisz et al.'s feature-set 0.7943.

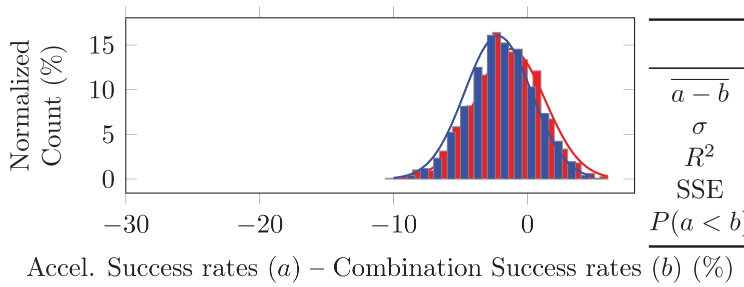
Implications of the results of this analysis will be discussed in the discussion section (section 5.4).

Each *source* vs the combination of the three *sources* for the 6 monitor setup

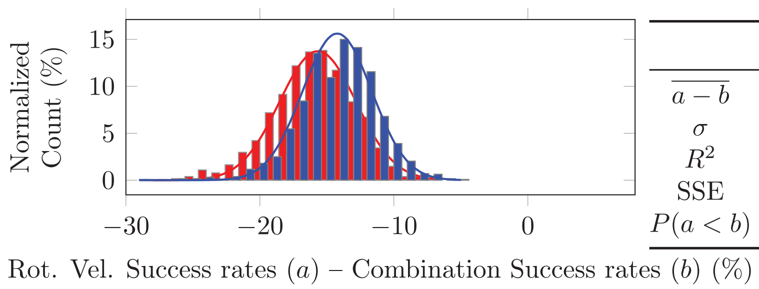
Success-rates obtained using feature-vectors extracted the combination of the three *sources* for the 6 monitor setup are shown in figure 5.1. Success-rates obtained using feature-vectors extracted from each of the three *sources* for the 6 monitor setup are shown in figure 5.6.

The differences of the success-rates obtained from each *source* from those obtained from the combination of the *sources* (*Bao-6* and *Kwapisz-6*) for the 6 monitor setup are summarised in figure 5.12.

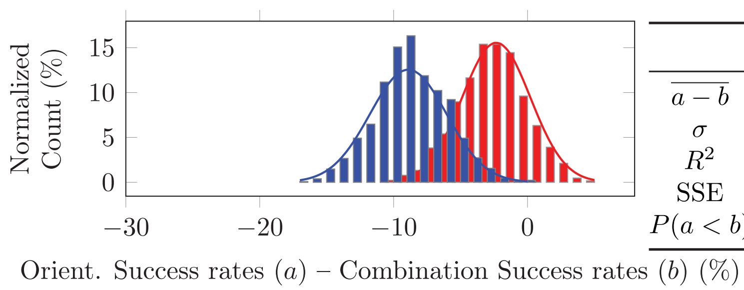
Figure 5.11 shows the distribution of the differences between the success-rates obtained from feature-vectors extracted from a combination of all three *sources* from



	Bao	***Kwapisz***
$\overline{a-b}$	-1.29	-2.20
σ	1.9238	1.8814
R^2	0.9916	0.9951
SSE	4.30	2.49
$P(a<b)$	0.7487	0.8789



	Bao	***Kwapisz***
$\overline{a-b}$	-15.67	-14.23
σ	2.0120	1.9133
R^2	0.9913	0.9879
SSE	4.37	7.31
$P(a<b)$	~ 1.0	~ 1.0



	Bao	***Kwapisz***
$\overline{a-b}$	-2.33	-8.90
σ	1.9003	2.0111
R^2	0.9917	0.9899
SSE	3.91	5.05
$P(a<b)$	0.8899	~ 1.0

Figure 5.12: Histograms of differences in success-rates obtained from feature-vectors extracted from each *source* from those obtained from feature-vectors extracted from all three *sources* of the 6 monitor setup. Differences in accelerometer success-rates (upper), rotational velocities (middle) and orientations (lower) are shown.

those of feature-vectors extracted from accelerations only, rotational velocities only and orientations only for the 6 monitor setup and for both feature-sets studied.

On average for the 6 monitor setup, feature-vectors from accelerations only result in success-rates that are 1.29% and 2.20% lower than feature-vectors from all three *sources* combined for Bao and Intille's feature-set and Kwapisz et al.'s feature-sets. Feature-vectors from rotational velocities result in success-rates that are 15.67% and 14.23% lower than feature-vectors from all three *sources* combined. Feature-vectors from orientations result in success-rates that are 2.33% and 8.90% lower than feature-vectors from all three *sources* combined. It should be noted that the mean differences in success-rates obtained for the 6 monitor setup are similar to those of the 3 monitor setup.

From the fitted Gaussian models, the probability of feature-vectors extracted from rotational velocities resulting in lower success-rates than feature-vectors extracted from the three *sources* combined, is observed to be almost certain for both feature-sets.

The probability of feature-vectors extracted from orientations resulting in lower success-rates than feature-vectors extracted from the three *sources* combined is almost certain for Kwapisz et al.'s feature-set but is 0.8899 for Bao and Intille's feature-set.

The probability of feature-vectors extracted from accelerations resulting in lower success-rates than feature-vectors extracted from the three *sources* combined is lower than any of the other *sources* is 0.7487 for Bao and Intille's feature-set and 0.8789 for Kwapisz et al.'s feature-set.

Discussion

In this section, the success-rates obtained by using each *source* was compared to the success-rates obtained by each other *source* and to the success-rates obtained while using the combination of the three *sources*, for both studied feature-sets.

Due to the results of the analysis performed in section 5.3 that found that the difference between the 3 monitor setup and the 6 monitor setup have a significant impact on the results obtained, the analysis is repeated for both the 3 monitor setup and the 6 monitor setup.

The difference in the result sets obtained were found to be statistically significant. Using the distributions of success-rates from different *sources*, accelerations were found to have the highest mean success-rates of the three *sources* for both feature-sets studied and for both monitor setups. Orientations were found to be second highest and rotational velocities found to have the lowest mean success-rates of the three *sources* for both feature-sets studied and for both monitor setups. However, overlaps in result sets from different *sources* but from the same feature-sets and monitors were observed.

Hence, the differences in success-rates between each of the three *sources* and those of each other *source* were studied. The differences between success-rates obtained from accelerations and those obtained from rotational velocities were 14.05%, 14.44%, 12.56% and 12.01% for Bao and Intille's feature-set using the 3 monitor setup and 6 monitor setup, and Kwapisz et al.'s feature-set using the 3 monitor setup and 6 monitor setup, respectively. In the same order, the differences between success-rates obtained from accelerations and those obtained from orientations were 3.01%, 1.13%, 8.26% and 6.72%, and those betweeen success-rates obtained from rotational velocities and those obtained from orientations were 11.07%, 13.32%, 4.44% and 5.18%.

The probability of success-rates obtained from accelerations being lower than those obtained from rotational velocities was found to be very low for both feature-sets and for both monitor setups. Similarly, the probability of success-rates obtained from orientations being lower than those obtained from rotational velocities was found to be very low for both feature-sets and for both monitor setups. The probability of success-rates obtained from accelerations being lower than those obtained from orientation was found to be very low for Kwapisz et al.'s feature-set for both monitor setups. Those of Bao and Intille's feature-set were slightly higher at 5.49% and 27.81% for the 3 monitor setup and 6 monitor setup respectively, but still show that the probability of accelerations having a higher success-rate than orientations is high.

Hence, we can conclude that of the three *sources*, accelerations result in highest activity recognition success-rates, while rotational velocities result in the lowest activity recognition success-rates.

Next, the differences in success-rates between each of the three *sources* and the combination of the three *sources* was studied. The differences were earlier in the section found to be statistically significant.

The differences between success-rates obtained from feature-vectors extracted from accelerations alone were found to have the lowest probability of being lower than those obtained from feature-vectors extracted from all three *sources* combined. In addition, the mean differences between success-rates obtained from feature-vectors extracted from accelerations and those obtained from feature-vectors extracted from all three *sources* were found to be the highest of the three *sources*. This was observed for both feature-sets and for both the 3 monitor setup and the 6 monitor setup.

The differences between success-rates obtained from feature-vectors extracted from rotational velocities alone were found to almost certainly be lower than those obtained from feature-vectors extracted from all three *sources* combined. In addition, the mean differences between success-rates obtained from feature-vectors extracted from rotational velocities and those obtained from feature-vectors extracted from all three *sources* were found to be the highest of the three *sources*. This was observed for both feature-sets and for both the 3 monitor setup and the 6 monitor setup.

The differences between success-rates obtained from feature-vectors extracted from orientations alone were found to almost certainly be lower than those obtained from feature-vectors extracted from all three *sources* combined for Kwapisz et al.'s feature-set for both the 3 monitor setup and the 6 monitor setup. The probability of success-rates in *Bao-Orient-3* being lower than those in *Bao-3* is 0.9804, which is higher than the probability of success-rates in *Bao-Orient-6* being lower than those in *Bao-6*.

The mean differences between success-rates obtained from feature-vectors extracted from orientations and those obtained from feature-vectors extracted from all three *sources* were found to be higher for Kwapisz et al.'s feature-set than for Bao and Intille's feature-set for both the 3 monitor setup and the 6 monitor setup.

However, the average difference between success-rates obtained from acceleration and those obtained from the combination of the three *sources* was found to be small (1%-2% depending on the feature-set). The average difference between success-rates obtained from orientations and those obtained from the combination of the three *sources* was found to be 2% while using Bao and Intille's feature-set but 9% while using Kwapisz et al.'s feature-set.

Hence for the two feature-sets studied and the activities gathered within this thesis, we can conclude that of the three *sources* studied, feature-vectors extracted from accelerations have success-rates that have the smallest average difference from those of feature-vectors extracted from all three *sources* and they have the lowest chance of being lower. Feature-vectors extracted from rotational velocities have success-rates that have the greatest mean difference from those of feature-vectors extracted from all three *sources* and they are almost certainly lower. Feature-vectors extracted from orientations have success-rates that are either highly likely or almost certainly lower than success-rates obtained from feature-vectors extracted from all three *sources* depending on the feature-set used. Kwapisz et al's feature-set results in success-rates that are almost certainly lower than those obtained from feature-vectors extracted from all three *sources* and having a higher mean difference, while Bao and Intille's feature-set results in success-rates that are sometimes higher than those obtained from feature-vectors extracted from all three *sources* and have a lower mean difference.

As it has been observed in this section, both accelerations and orientations result in higher success-rates than rotational velocities for the activities and feature-sets studied. One common attribute between accelerations and orientations that rotational velocities do not share is that both accelerations and orientations include the global pitch of the monitor. Accelerations do so by encoding within the signal the direction of gravity (in the form of the gravity vector) and hence encode the pitch of the monitor. Orientations encode within the signal the pitch, bearing and roll of the monitor by definition.

Rotational velocities, however, are the changes in orientation. Hence, when a subject is performing an activity that results in some of the monitors mounted on him being stationary, the gyroscopes from the monitors would read values close to zero. Accelerations and orientations, however, would read values that include in them the pitch of the monitor within that stationary period.

A possible explanation for the higher performance of accelerations and orientations compared to rotational velocities is that this additional information may be helping to distinguish activities where some monitors mounted on the subject's body are stationary but in different orientations than other activities. This would result in higher success-rates than when that information is not available. An example is standing compared to sitting using a thigh mounted monitor, or sitting up while working on a PC or writing on a desk compared to sitting back while watching TV by using a chest mounted monitor.

The implications of the results obtained in this section on wearable activity recognition systems are that, in a situations where attaining the highest success-rates is important, if all three *sources* can be obtained and processed, then combining all three is likely to result in a higher success-rate than any one individual *source*. However, if this is not possible, then accelerations should be prefered to orientations, which in turn should be preferred to rotational velocities.

5.5 Performance comparison of body-locations

In this section we are interested in comparing the performance of the body-locations based on the activity classification success-rate of the data captured by the monitors mounted on those body-locations.

The research question asked is in this section is: is there a significant difference in activity classification rates when a monitor is mounted on different body-locations? If so, which location yields the highest activity classification accuracy?

Knowing which body-locations perform better than others allows activity recognition researchers to select locations for monitoring activities that would recognise activities most accurately.

Section 5.6 will compare the activity classification performance of combinations of body-locations in order to determine which combinations of body-locations result in better performance than others. This section focusses on individual body-locations only.

Methodology

In order to answer the research question, success-rates computed from data captured by each monitor were compared to success-rates obtained from data captured by other monitors.

The data gathered includes activities with 6 monitors mounted on each subject and also activities with 3 monitors mounted on each subject. The activities with only 3 monitors mounted are walking and running only. The rest of the activities were gathered using 6 monitors. To be able to determine which body-locations results in the higher success-rates than other body-locations, all 6 monitors were used and hence only activities recorded with all 6 monitors were used.

Analysis performed in section 5.3 found that the difference between the 3 monitor setup and the 6 monitor setup in terms of the success-rates was statistically significant but at most a mean difference of 4% was observed. Analysis performed in section 5.4 found the same trends in success-rates obtained from accelerations, rotational velocities and orientations between the 3 monitor setup and the 6 monitor setup.

The exclusion of the walking and running activities is not expected to significantly impact the the trends observed in the analysis and hence not impact the final conclusions. This is because, walking and running are whole body activities, and hence should be detected well on any of the arms, legs or torso. This was observed for walking by Kern, Schiele, and Schmidt (2003), and walking and running by Henpraserttae et al. (2011).

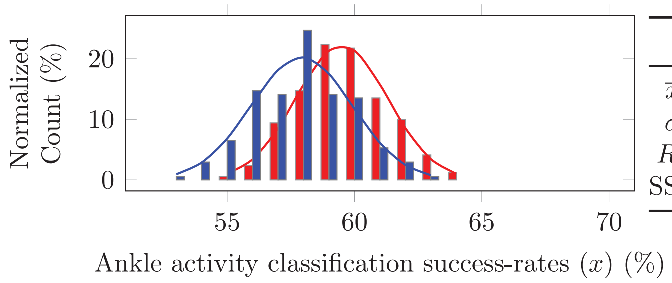
Hence, the analysis in this section was performed for the two feature-sets using six monitors but not including walking and running.

12 sets of results were computed using algorithm 6. The parameters provided to algorithm 6 for the result sets are given in table 5.6.

To find which body-locations result in higher success-rates than other body-locations, the differences in every pair of the result sets of each feature-set was computed. For each pair, a Gaussian model was then fitted over the differences using maximum-likelihood estimation and the mean and standard deviation of the model extracted. The mean and standard deviation of the fitted model are then used to estimate the likelihood of data captured from one body-location having a higher success-rate than data captured from another body-location.

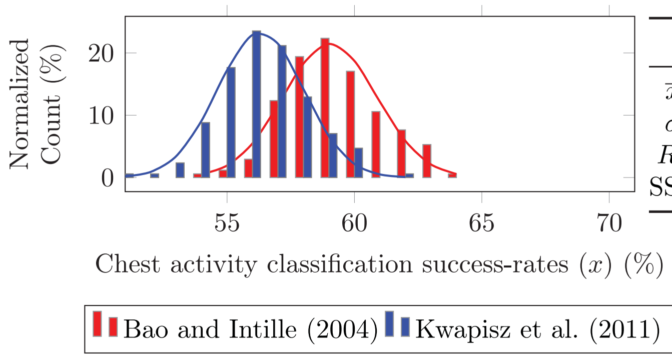
Table 5.6: Parameters provided to algorithm 6 so as to compare the activity classification success-rates obtained from data captured by each monitor for the two feature-sets studied.

Result Set	***FeatureSet***	***Activities***	***Sources***			***Monitors***					
			Accel.	**Rot. Vel.**	**Orient.**	**A**	**C**	**DUA**	**DW**	**NDW**	**T**
Bao-Ankle	Bao and Intille (2004)	All except run & walk	✓	✓	✓	✓					
Bao-Chest	Bao and Intille (2004)	All except run & walk	✓	✓	✓		✓				
Bao-DUArm	Bao and Intille (2004)	All except run & walk	✓	✓	✓			✓			
Bao-DWrist	Bao and Intille (2004)	All except run & walk	✓	✓	✓				✓		
Bao-NDWrist	Bao and Intille (2004)	All except run & walk	✓	✓	✓					✓	
Bao-Thigh	Bao and Intille (2004)	All except run & walk	✓	✓	✓						✓
Kwapisz-Ankle	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓	✓					
Kwapisz-Chest	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓		✓				
Kwapisz-DUArm	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓			✓			
Kwapisz-DWrist	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓				✓		
Kwapisz-NDWrist	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓					✓	
Kwapisz-Thigh	Kwapisz et al. (2011)	All except run & walk	✓	✓	✓						✓



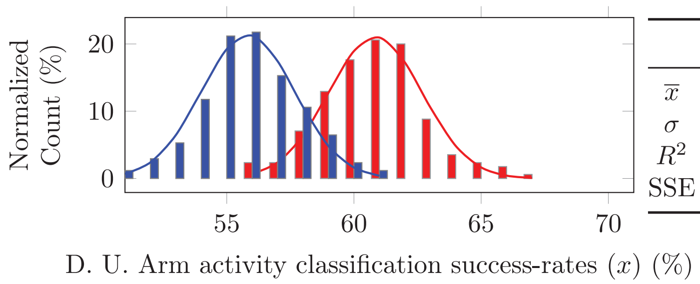
	Bao	***Kwapisz***
$\overline{x}$	59.50	57.93
σ	1.6001	1.6764
R^2	0.9798	0.9039
SSE	11.7	56.3

Figure 5.13: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle.



	Bao	***Kwapisz***
$\overline{x}$	59.04	56.31
σ	1.6134	1.5469
R^2	0.9546	0.9854
SSE	28.1	10.5

Figure 5.14: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Chest.

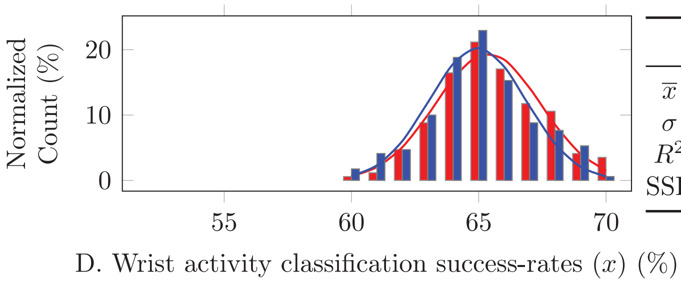


	Bao	***Kwapisz***
$\overline{x}$	60.84	55.84
σ	1.6141	1.6105
R^2	0.9681	0.9729
SSE	20.0	15.8

Figure 5.15: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. U. Arm.

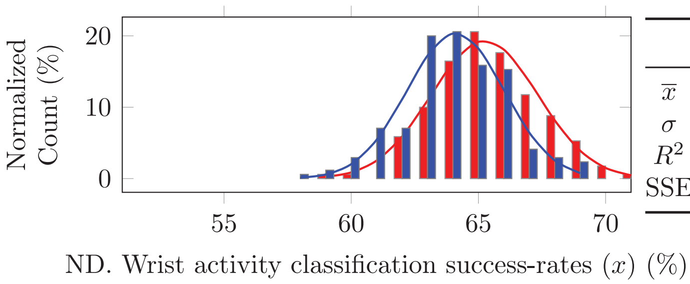
Results

The result sets obtained using algorithm 6 and the parameters given in table 5.6 are summarised in figure 5.13 (Ankle), figure 5.14 (Chest), figure 5.15 (Dominant Upper Arm), figure 5.16 (Dominant Wrist), figure 5.17 (Non-dominant Wrist) and figure 5.18 (Thigh/Phone). The mean success-rates of the result sets obtained are given in table 5.7.



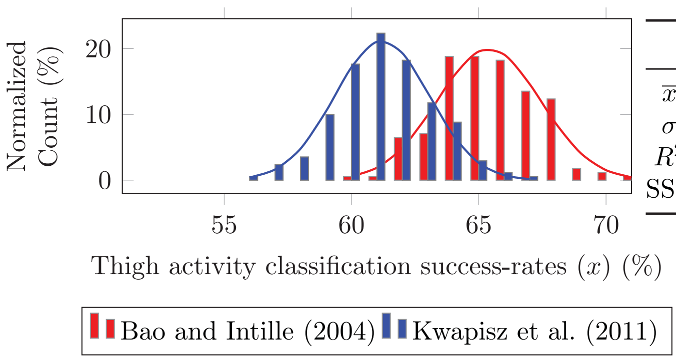
	Bao	***Kwapisz***
$\overline{x}$	65.36	64.95
σ	1.7053	1.6388
R^2	0.9516	0.9154
SSE	23.3	43.3

Figure 5.16: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. Wrist.



	Bao	***Kwapisz***
$\overline{x}$	65.24	64.10
σ	1.7251	1.6437
R^2	0.9835	0.9191
SSE	9.1	49.9

Figure 5.17: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the ND. Wrist.



	Bao	***Kwapisz***
$\overline{x}$	65.40	61.20
σ	1.6941	1.6211
R^2	0.9329	0.9814
SSE	42.9	12.1

Figure 5.18: Distribution of success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Thigh.

Table 5.7: Mean success-rates of result sets computed from data gathered from each body-location.

Result Set	Bao and Intille (2004)	Kwapisz et al. (2011)
Ankle	59.5%	57.9%
Chest	59.0%	56.3%
Dominant upper arm	60.8%	55.9%
Dominant wrist	65.4%	65.0%
Non-dominant wrist	65.2%	64.1%
Thigh	65.4%	61.2%

From the figures, we can observe that the mean success-rates fit Gaussian distributions and are close to each other, with their means ranging from 55% to 65%.

A paired two-sample two-tailed t-test with $\alpha = 0.05$ between each pair of result sets was run to test the null hypothesis that the pairs of result sets came from independent random samples from normal distributions with equal means and unknown variances. The test did not assume that the two result sets had equal variance by using Satterthwaite's approximation of the effective degrees of freedom.

The tests rejected the null hypothesis for all pairs of result sets except: *Bao-Ankle* and *Bao-Chest*, *Bao-DWrist* and *Bao-NDWrist*, *Bao-DWrist* and *Bao-Thigh*, and *Bao-NDWrist* and *Bao-Thigh*. Hence, for the pairs of result sets named, there is insufficient evidence to support the hypothesis that the result sets in the pairs differ. The named result set pairs can be summarised into two groups of result sets that are similar to each other: result sets computed from data captured from the ankle and chest, and those computed from data captured from the dominant wrist, non-dominant wrist and thigh (or phone). Both groups of result sets were computed using Bao and Intille's feature-set.

The differences in all result sets computed using Kwapisz et al.'s feature-set were found to be statistically significant.

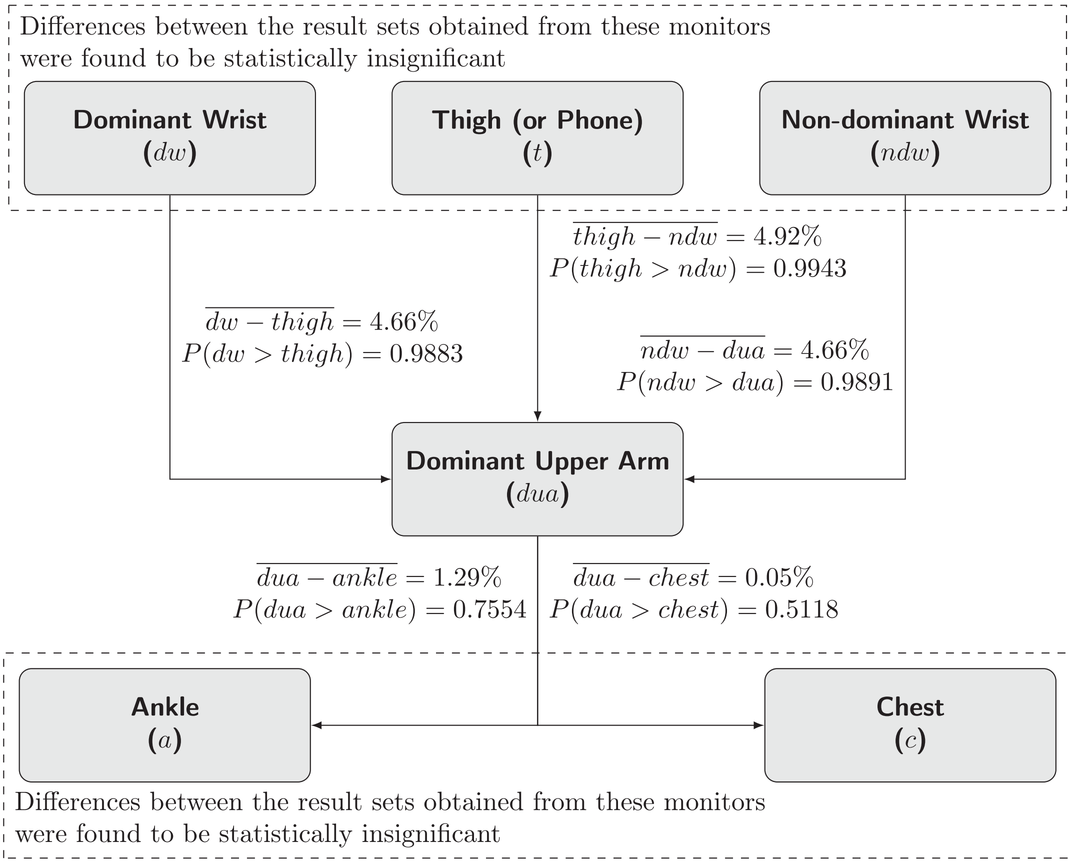


Figure 5.19: Illustration showing the ranking of body-locations on which monitors were mounted, from highest performing to lowest performing, based on the probability of obtaining higher activity classification success-rates from data of one body-locations than that of the other body-locations. The results are generated from data extracted using Bao and Intille's feature-set.

Illustrations of the distributions of differences in the result sets are given in appendix C. A summary of the distributions of the differences of result sets computed from data captured from different body-locations for the two feature-sets studied is given in table 5.8.

After sorting the result sets, such that result sets that are more likely to have success-rates that are higher than the equivalent success-rates in other result sets, the resulting relationships between the result sets are given in figure 5.19 for Bao and Intille's feature-set and figure 5.20 for Kwapisz et al.'s feature-set.

As illustrated in figure 5.19, three groups of body-locations were found based on the comparison of the performance of their captured data using Bao and In-

Table 5.8: Summary of distributions of differences in result sets computed from data captured from the six monitors for the two feature-sets studied. The columns represent the result sets compared, the mean of the signed differences in the two result sets, and the probability of the first result set having a higher success-rate than the second result set. Illustrations of the distributions of differences in the result sets are given in appendix C.

Result Set a	**Result Set b**	$\overline{a-b}$	$p(a>b)$
Bao-Ankle	*Bao-Chest*	0.05	0.5118
Bao-Ankle	*Bao-DUArm*	−1.29	0.2446
Bao-Ankle	*Bao-DWrist*	−5.71	0.0027
Bao-Ankle	*Bao-NDWrist*	−5.95	0.0015
Bao-Ankle	*Bao-Thigh*	−5.93	0.0010
Bao-Chest	*Bao-DUArm*	−1.70	0.1936
Bao-Chest	*Bao-DWrist*	−6.35	0.0009
Bao-Chest	*Bao-NDWrist*	−6.17	0.0004
Bao-Chest	*Bao-Thigh*	−6.56	0.0005
Bao-DUArm	*Bao-DWrist*	−4.66	0.0117
Bao-DUArm	*Bao-NDWrist*	−4.66	0.0109
Bao-DUArm	*Bao-Thigh*	−4.92	0.0057
Bao-DWrist	*Bao-NDWrist*	0.30	0.5536
Bao-DWrist	*Bao-Thigh*	0.23	0.5447
Bao-NDWrist	*Bao-Thigh*	−0.21	0.4591
Kwa-Ankle	*Kwa-Chest*	1.26	0.7477
Kwa-Ankle	*Kwa-DUArm*	2.22	0.8907
Kwa-Ankle	*Kwa-DWrist*	−7.35	0.0001
Kwa-Ankle	*Kwa-NDWrist*	−6.33	0.0009
Kwa-Ankle	*Kwa-Thigh*	−3.22	0.0544
Kwa-Chest	*Kwa-DUArm*	0.44	0.5904
Kwa-Chest	*Kwa-DWrist*	−8.72	0.0000
Kwa-Chest	*Kwa-NDWrist*	−7.67	0.0001
Kwa-Chest	*Kwa-Thigh*	−4.70	0.0113
Kwa-DUArm	*Kwa-DWrist*	−9.42	0.0000
Kwa-DUArm	*Kwa-NDWrist*	−8.22	0.0000
Kwa-DUArm	*Kwa-Thigh*	−5.50	0.0043
Kwa-DWrist	*Kwa-NDWrist*	0.84	0.6647
Kwa-DWrist	*Kwa-Thigh*	3.78	0.9683
Kwa-NDWrist	*Kwa-Thigh*	2.84	0.9311

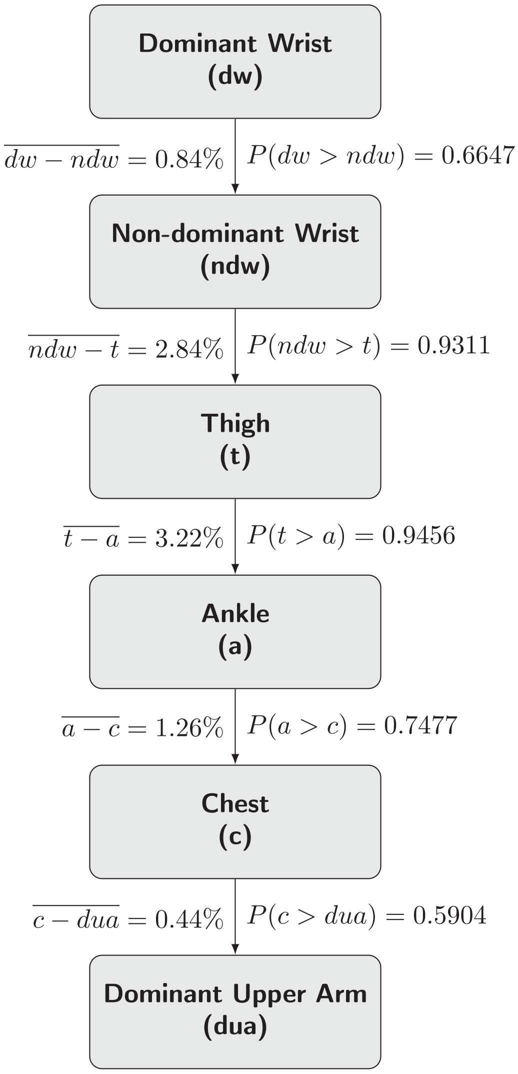


Figure 5.20: Illustration showing the ranking of body-locations on which monitors were mounted, from highest performing to lowest performing, based on the probability of obtaining higher activity classification success-rates from data of one body-locations than that of the other body-locations. The results are generated from data extracted using Kwapisz et al.'s feature-set.

tille's feature-set: the dominant wrist, thigh (phone) and non-dominant wrist; the dominant upper arm; and the ankle and chest. Data captured from each of the body-locations in the first group was found to be more likely to result in higher success-rates than the data captured from the dominant upper arm. Similarly, data captured from the dominant upper arm was found to be likely to result in higher success-rates than the data captured from the ankle or chest.

While using Kwapisz et al.'s feature-set, data captured from the dominant wrist, non-dominant wrist, and thigh was still found to be more likely to result in higher success-rates than the data captured from the dominant upper arm. However, unlike what was observed with Bao and Intille's feature-set, data captured from the ankle and chest was also found to be more likely to result in higher success-rates than the data captured from the dominant upper arm.

Discussion

In this section, the success-rates obtained from data captured from each monitor(body location) was compared to success-rates obtained from data captured from each other monitor(alternate body location) for both feature-sets studied.

Data from all six monitors was used, but data from walking and running activities was excluded because these activities were gathered with three monitors only. Excluding walking and running means that the results in this section do not include any impact of walking and running on the activity classification. For example, if the data captured from some of the body-locations could recognise walking and running well while data captured from other body-locations would not, then the inclusion of walking and running would skew the results against some of the monitors and in favour of the other monitors. However, this is unlikely since walking and running are whole body activities and have been recognised with high success-rates in different body-locations (Henpraserttae et al., 2011). In addition, section 5.4 observed similar trends in activity classification of data with and without walking and running.

The difference in the results sets were found to be statistically significant for all result sets computed using Kwapisz et al.'s feature-sets, but found differences in some of result sets computed from Bao and Intille's feature-set not statistically significant. The result sets whose differences were found not statistically significant were those computed from: the dominant wrist, thigh and non-dominant wrist; and the ankle and chest.

Gaussian models were then fitted on the signed differences between the result sets, and the mean and standard deviation of the models was obtained. Using the fitted models, the likelihood of data captured from one body-location having a higher success-rate than data captured from another body-location was estimated. The probabilities were then used to sort the result sets.

From the analysis, it was observed that success-rates computed from data captured from the dominant wrist, thigh (or phone) and non-dominant wrist are more likely to be higher than success-rates computed from data captured from the chest, dominant upper arm and ankle, for both feature-sets studied. While differences between result sets computed from data captured by the dominant wrist, thigh and non-dominant wrist were found to be statistically insignificant using Bao and Intille's feature-set, using Kwapisz et al.'s feature-set found that success-rates computed from data captured from the dominant wrist were more likely to be higher than those computed from data captured from the non-dominant wrist, which in turn were more likely to be higher than success-rates computed from data captured from the thigh (or phone).

However, success-rates computed from data captured from the dominant upper arm are likely to be higher than those computed from data captured from the chest or ankle using Bao and Intille's feature-set, but are likely to be lower while using Kwapisz et al.'s feature-set. Between the ankle and chest, it was found that success-rates computed from data captured from the ankle are likely to be higher than those computed from data captured from the chest using Kwapisz et al.'s feature-set.

It is possible that success-rates from data captured from the wrists are higher than those of other locations because many of the activities require the use of the hands to manipulate objects (knifes, dusters, brooms, vacuum cleaners, irons, clothes, pens, keyboards, mouses, etc.). Data from the motions used to manipulate the different objects while doing different activities possibly distinguishes the activities, or types of activities, that the subjects are doing and results in better activity classification success-rates.

Surprisingly, success-rates from data captured from the thigh were found be higher than those from data captured from the dominant upper arm, ankle and chest for both feature-sets. It is possible that this is a result of the thigh being able to distinguish sitting activities from standing/upright activities using the orientation of the sensor on the thigh.

The implications of the results obtained in this sections are that if any one location on the subject's body should be selected to mount sensors to recognise Activities of Daily Living, then the wrists should be preferred to the thigh, which in turn should be preferred to the dominant upper arm, ankle or chest.

5.6 Performance change with increase in number of body-locations monitored

In this section we are interested in analysing the change in success-rates with increase in number of monitors at different body-locations.

Hence, the research question in this section is: what is the relationship between activity classification success-rates as a function of the number of monitors mounted on different locations of the body?

Results of this section might differ from those obtained from section 5.5 since using data obtained from multiple body-locations can result in redundancy in the information that can distinguish activities from each other. Hence, for example, two body-locations that were individually highly ranked in section 5.5, do not necessarily result in higher success-rates when the data of the two body-locations is combined.

Methodology

The data gathered includes activities with 6 monitors mounted on each subject and also activities with 3 monitors mounted on each subject. The activities with only 3 monitors mounted are walking and running. The rest of the activities were gathered using 6 monitors. To be able to determine the success-rate obtained using data captured from different sets of body-locations, all 6 monitors were used and hence only activities recorded with all 6 monitors were used.

Analysis performed in section 5.3 found that the difference between the 3 monitor setup and the 6 monitor setup in terms of the success-rates was statistically significant but at most a mean difference of 4% was observed. Analysis performed in section 5.4 found the same trends in success-rates obtained from accelerations, rotational velocities and orientations between the 3 monitor setup and the 6 monitor setup.

As explained in section 5.5, the exclusion of the walking and running activities is not expected to significantly impact the trends observed in the analysis and hence not impact the final conclusions. This is because, walking and running are whole body activities, and hence should be detected well on arms, legs or the torso.

Hence, analysis in this section was performed for the two feature-sets using six monitors but not including walking and running.

To answer the research question, the highest average success-rates obtained from data captured from sets of one to six body-locations was computed. That is, algorithm 7 was called for values of N in the range 1 to 6 with increments of 1, and for both Bao and Intille's and Kwapisz et al.'s feature-sets. N represents the number of body-locations used for activity classification.

Algorithm 7 Find the highest success-rate obtained using data captured from N body-locations, using all activities except for walking and running, and all three *sources*: accelerations, rotational velocities and orientations.

```
procedure TestActivityRecognition(FeatureSet, N)
    HighestSuccessRate ← 0                ▷ Highest success-rate obtained so far.
    HighestMonitorGroup ← NULL            ▷ Monitor group with the highest success-rate.
    BestResultSet ← NULL                  ▷ Result set with the highest success-rate.
    BestMonitorGroups ← { }               ▷ Set of monitor groups that have statistically in-
                                            significant differences with the group with the
                                            highest success-rate.
    AllMonitors ← get all six monitors
    for all monitor_group ∈ all sets of N monitors from AllMonitors do
        Call algorithm 6 with parameters:
            1. Feature set: FeatureSet
            2. Activities: All activities excluding walking and running.
            3. Sources: Accelerations, rotational velocities and orientations.
            4. Monitors: monitor_group
        Obtain result set res.
        Fit Gaussian model onto result set and obtain fitted mean m.
        if m > HighestSuccessRate then
            HighestSuccessRate ← m
            BestResultSet ← res
            HighestMonitorGroup ← monitor_group
            BestMonitorGroups ← { }       ▷ Clear set of best monitor groups.
        else
            if TTest(BestResultSet, res) > 0.05 then
                                          ▷ Check whether differences between res and
                                            BestResultSet are statistically significant.
                BestMonitorGroups.add(monitor_group)
            end if
        end if
    end for
    return HighestSuccessRate, HighestMonitorGroup, BestMonitorGroups
end procedure
```

Results

Results obtained from executing algorithm 7 with $N = [1, 6]$ in intervals of 1 are shown in table 5.9 for Bao and Intille's feature-set, and table 5.10 for Kwapisz et

Table 5.9: The sets of body-locations resulting in the highest success-rate for every set of N body-locations along with the mean and standard deviation of the success-rates obtained. The data was processed using Bao and Intille's feature-set.

N	Body-locations						Mean	Standard Deviation
	A	C	DUA	DW	NDW	T		
1				✓			65.37	1.7051
1					✓		65.29	1.7349
1						✓	65.36	1.6845
2					✓	✓	75.61	1.5995
3				✓	✓	✓	79.08	1.6142
4		✓		✓	✓	✓	79.10	1.7238
5	✓	✓		✓	✓	✓	79.21	1.6640
5		✓	✓	✓	✓	✓	79.35	1.5331
6	✓	✓	✓	✓	✓	✓	79.00	1.5153

Legend
A: Ankle, C: Chest, DUA: Dominant Upper Arm
T:Thigh DW: Dominant Wrist, NDW:Non-dominant Wrist

al.'s feature-set.

From table 5.9, it can be observed that for $N = 1$ three different body-locations result in the success-rates that are statistically similar. The highest success-rate for $N = 2$ is observed to result from the non-dominant wrist and thigh. The highest success-rate for $N = 3$ is observed to result from sets that include the wrists and the thigh. Next, for $N = 4$ the chest is included into the best set of $N = 3$, and for $N = 5$ either the ankle or the dominant upper arm is included into the best set of $N = 4$. These observations from the results are illustrated in figure 5.21.

The two wrists and the thigh form the three body-locations that result in the highest success-rates, either individually or when combined. This is similar to what was observed in section 5.5 for the Bao and Intille's feature-set.

Section 5.5 found that data captured from the dominant upper arm results in higher success-rates than either the data from the chest or the ankle while using

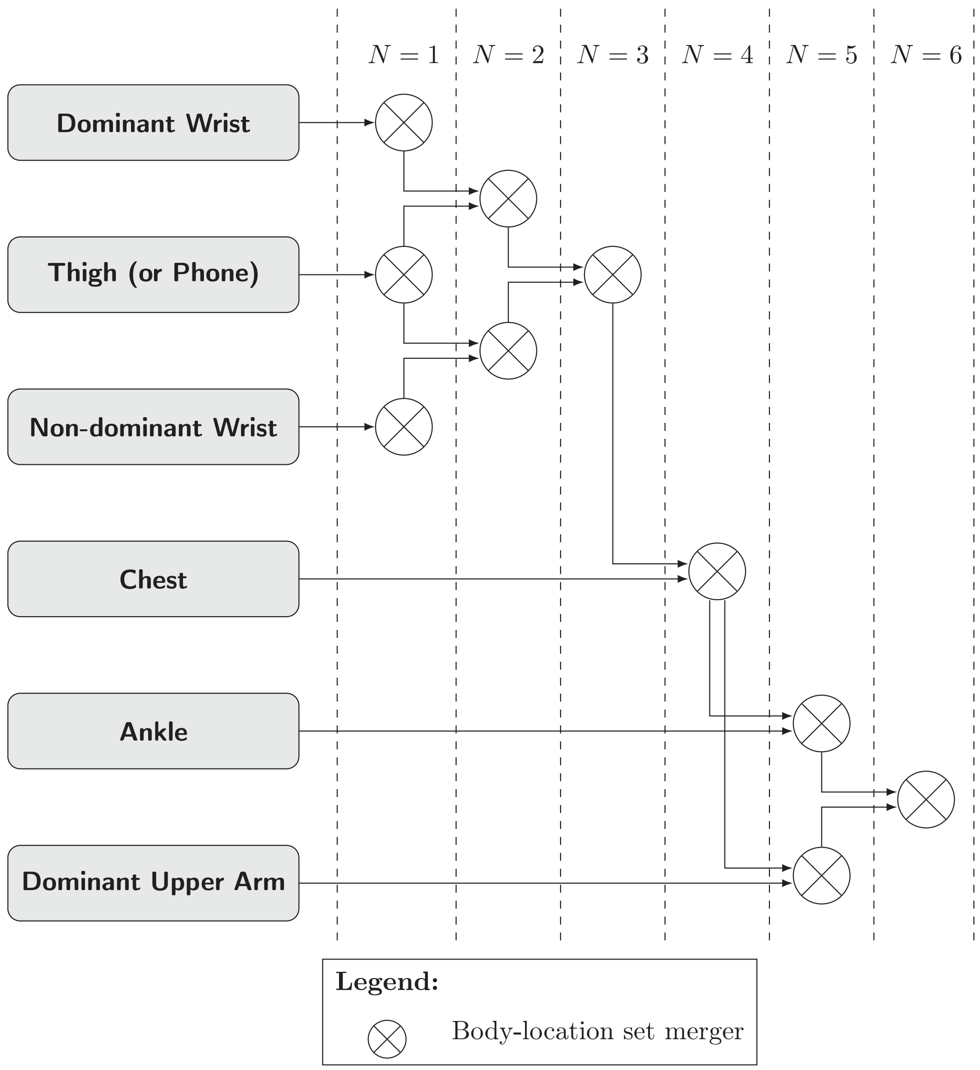


Figure 5.21: Illustration sets of body-locations, whose captured data result in the highest success-rates for the given value of N. At each value of N, the body-location set mergers in the column represent the sets of body-locations that result in the highest success-rate. Every body-location set merger merges two sets of body-locations, or a set of body-locations and a body-location, or is a set of one body-location. The illustration is based on results obtained from Bao and Intille's feature-set and shown in table 5.9.

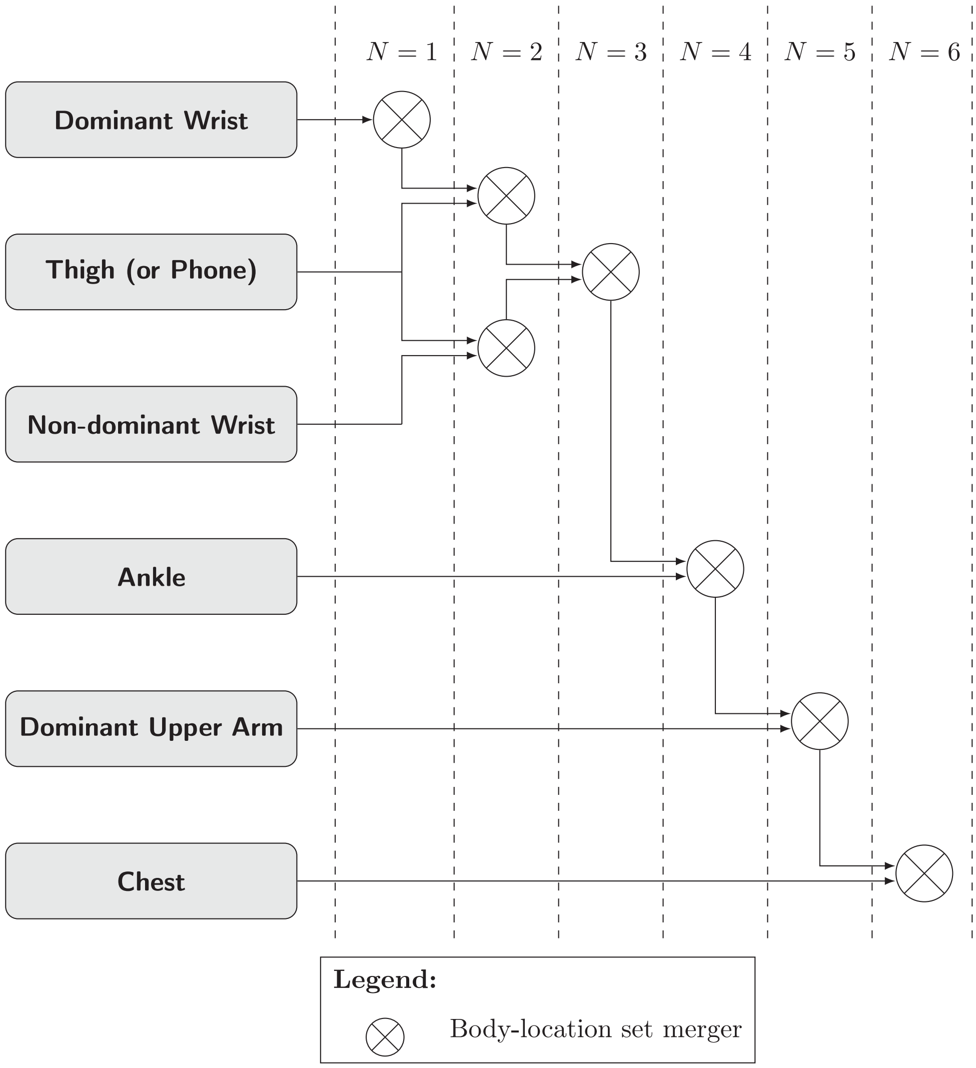


Figure 5.22: Illustration sets of body-locations, whose captured data result in the highest success-rates for the given value of N. At each value of N, the body-location set mergers in the column represent the sets of body-locations that result in the highest success-rate. Every body-location set merger merges two sets of body-locations, or a set of body-locations and a body-location, or is a set of one body-location. The illustration is based on results obtained from Kwapisz et al.'s feature-set and shown in table 5.10.

Table 5.10: The sets of body-locations resulting in the highest success-rate for every set of N body-locations along with the mean and standard deviation of the success-rates obtained. The data was processed using Kwapisz et al.'s feature-set.

N	Body-locations						Mean	Standard Deviation
	A	C	DUA	DW	NDW	T		
1				✓			64.95	1.6402
2				✓		✓	74.15	1.7056
2					✓	✓	73.86	1.6280
3				✓	✓	✓	77.61	1.6831
4	✓			✓	✓	✓	77.95	1.7267
5	✓		✓	✓	✓	✓	78.01	1.5764
6	✓	✓	✓	✓	✓	✓	77.23	1.6552

Legend
A: Ankle, C: Chest, DUA: Dominant Upper Arm
T:Thigh DW: Dominant Wrist, NDW:Non-dominant Wrist

Bao and Intille's feature-set. However, the results given in table 5.9 show that, for $N = 4$, the set of monitors mounted on the chest, wrists and thigh result in higher success-rates than either the sets of monitors mounted on ankle, wrists and thigh or dominant upper arm, wrists and thigh. For $N = 5$, both the set that includes the dominant upper arm and the set that includes the ankle are found to result in statistically similar success-rates.

Kwapisz et al.'s results, as shown in table 5.10, show that data captured from the dominant wrist results in the highest success-rates for $N = 1$. Either wrist and the thigh result in the highest success-rates when $N = 2$ and both wrists and the thigh when $N = 3$. Next, the ankle location is included for $N = 4$, dominant upper arm for $N = 5$ and finally the chest is included for $N = 6$. These observations from the results are illustrated in figure 5.22.

Data captured from the dominant wrist, followed by that captured from the non-dominant wrist, thigh (or phone), and ankle result in the highest success-rates in that order. This is similar to observations made in section 5.5 for Kwapisz et al.'s feature-set.

Data from the dominant upper arm was prefered over data from the chest for $N = 5$. However, in section 5.5, it was observed that data captured from the chest was found to result in higher success-rates than data captured from dominant upper arm while using Kwapisz et al.'s feature-set.

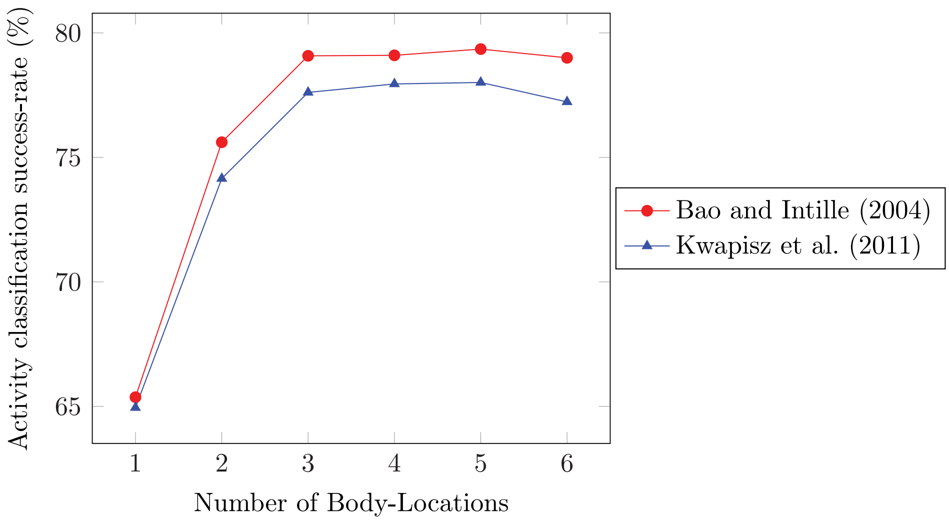


Figure 5.23: Success-rate as a function of the number of body-locations used for data capture for the two feature-sets studied.

Figure 5.23 shows a plot of the maximum success-rates obtained as a function of N. For both feature-sets studied, a rapid increase in success-rates is observed within the range $N = [1, 3]$. The success-rates then remain relatively constant in the range $N = [4, 6]$.

Although only six points exist per feature-set, the maximum success-rates as a function of N appear to fit the sum of a diminishing returns relationship and a constant. We can model this relationship using equation 5.1.

$$y' = C - A\, \mathrm{e}^{-Sx} \tag{5.1}$$

where, $y' \in [0, 100]$: estimated y value,

$C \in [0, 100]$: maximum estimated y value,

$$A \in [0, 100]\text{: } y \text{ scaling factor,}$$

$$S \in [1, \infty)\text{: } x \text{ scaling factor}$$

After fitting equation 5.1 on the data presented in figure 5.23 using the Trust-Region-Reflective Least-Squares regression algorithm, equation 5.2 and equation 5.3 are obtained for results obtained using Bao and Intille's and Kwapisz et al.'s feature-set respectively.

$$\text{Success rate}_{\ bao} \approx 79.35 - 56.73 \ \mathrm{e}^{-1.398N} \tag{5.2}$$

$$\text{Success rate}_{\ kwapisz} \approx 77.93 - 49.54 \ \mathrm{e}^{-1.336N} \tag{5.3}$$

$$\text{where, } N \in [1, \inf)$$

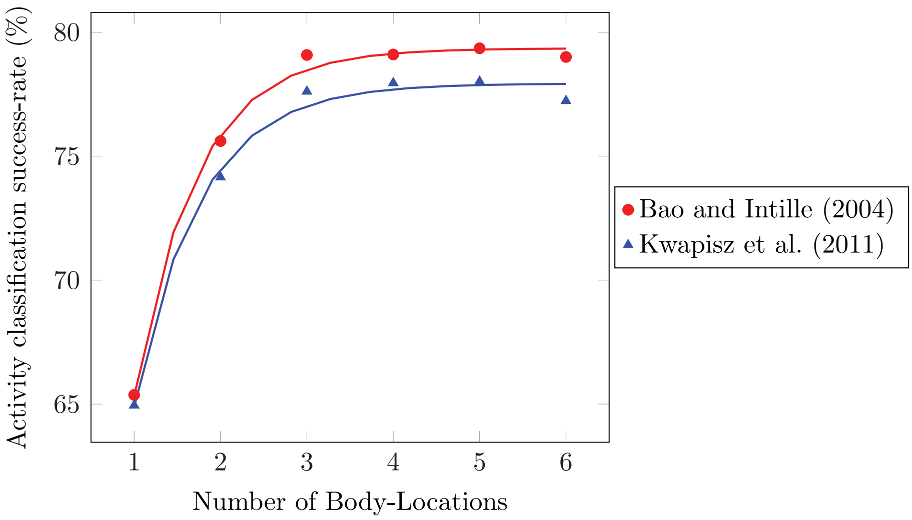


Figure 5.24: Success-rate as a function of the number of body-locations used for data capture for the two feature-sets studied with fitted equation 5.2 and equation 5.3.

Goodness of fit values are shown in table 5.11 and residuals are shown in figure 5.25. The goodness of fit values show that the equations fit the data well. However, the residuals of the two result sets are observed to be correlated. The correlation coefficient of the residuals is 0.9020. From that, we can infer that the model does not fully represent the pattern observed in the data and hence can be refined further. However, the residuals are small ($< 1\%$ relative error) hence the

Table 5.11: Goodness of fit values for the modelled success-rates as a function of number of body-locations for the two feature-sets studied. The equations fitted are equation 5.2 and equation 5.3, the observations are presented in figure 5.23.

Goodness of fit value	**Bao and Intille (2004)**	**Kwapisz et al. (2011)**
Sum of Square Error	0.5344	1.015
R^2	0.9965	0.9923
Adjusted R^2	0.9941	0.9871
Root Mean Square Error	0.4221	0.5817

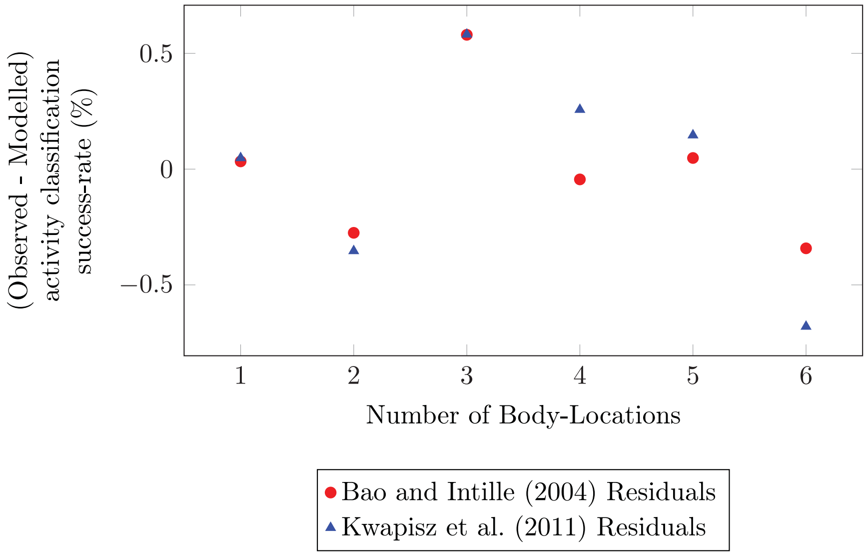


Figure 5.25: Residuals of fitted equation 5.2 and equation 5.3 on the observed success-rates shown in figure 5.23.

error introduced to the model is negligeable. The percentage relative error between the observed and estimated values is in the range $[\sim 0.0\%, 0.7\%]$ for Bao and Intille's feature-set, and $[\sim 0.0\%, 0.8\%]$ for Kwapisz et al.'s feature-set. The error-rates were computed using equation 5.4.

$$Error = \left| \frac{observed - estimated}{observed} \right| 100 \tag{5.4}$$

From the fitted equations, it can be observed that the diminishing returns relationship at most attains a maximum success-rate of 79.35% and 77.93%, and covers a range of 56.73% and 49.54%, for Bao and Intille's and Kwapisz et al.'s feature-set respectively.

5.7 Discussion

In this section, success-rates obtained from a set of N body-locations have been studied for $N = [1, 6]$.

Similar to section 5.5, data from all six monitors was used, but data from walking and running activities was excluded because these activities were gathered with three monitors only.

For each value of N, the set of body-locations, whose captured data resulted in the highest success-rate was computed. In addition, other sets of body-locations, whose captured data resulted in success-rates that are statistically similar to those of the set with the highest success-rate, were also computed.

From the results, it was observed that the top three body-locations, whose captured data resulted in the highest success-rates in section 5.5 were also the body-locations in the sets that resulted in the highest success-rates. These three body-locations are: the dominant wrist, the non-dominant wrist and the thigh.

Using Bao and Intille's feature-set, differences in the result sets computed from data captured from the three body-locations individually (i.e. with $N = 1$), were statistically insignificant, but the set of non-dominant wrist and thigh was selected for $N = 2$.

Using Kwapisz et al.'s feature-set, the dominant wrist was prefered for $N = 1$, but data captured from either wrist and thigh, was found to result in the highest success-rates for $N = 2$.

Patterns in results of $N = 4$ and $N = 5$ differ with those observed in section 5.5. Section 5.5 found that data gathered from the dominant upper arm resulted in higher success-rates than data gathered from either ankle or chest while using Bao and Intille's feature-set. However, in the results obtained from this section, data gathered from the chest was found to result in higher success-rates when combined with data captured from the wrists and thigh, than either data captured from the ankle or dominant upper arm. For $N = 5$, data captured from the dominant upper arm and the ankle was found to result in similar success-rates when combined with data from the chest, wrists and thigh.

Using Kwapisz et al.'s features, section 5.5 found that data captured from the chest resulted in higher success-rates than data captured from the dominant upper arm. For $N = 5$, results in this section found that, data captured from the dominant upper arm results in better success-rates than data captured from the chest when combined with data captured from the ankle, wrists and thigh.

When maximum success-rates were plotted as a function of N, it was observed that the data resembled a diminishing returns relationship. Equation 5.1 was then fitted onto the data, and the goodness of fit values given in table 5.11 showed that the equations fit the data well. From the plot of the residuals, it was observed that the residuals were highly correlated. This implies that the model does not fully represent the pattern observed in the data. However, the magnitude of the residuals was observed to be small and hence the relative error introduced is negligeable.

From the fitted models, it was observed that the maximum estimated success-rate is 79.35% and 77.93% using Bao and Intille's and Kwapisz et al.'s feature-set respectively. It was also observed that the diminishing returns relationship covers an estimated 56.73% and 49.54% using Bao and Intille's and Kwapisz et al.'s feature-set respectively. However, these success-rates could change if other activity classification parameters were changed, for example, window length, window overlap, the set of activities classified.

The observation that the relationship between success-rates and number of monitors used from different body-locations is that of a diminishing returns relationship is not surprising. As monitors are mounted on more body-locations, less and less additional unique information is made available for activity classification to make use of to identify the activity. The extreme case is where the monitors are overlapping each other, in which case we would expect no additional information from the additional sensor.

The results obtained in this section imply that, with the right placement, the first few monitors mounted on the subject's body contribute the most to the activity classification success-rate. Each additional monitor contributes lesser information than the preceeding monitor. The best locations for Activities of Daily Living are on the wrists, followed by the thigh, then either ankle, dominant upper arm or chest, depending on the feature-set used.

5.8 Conclusion

In this chapter, the activity classification performance stemming from the different feature-sets, *sources*, and body-locations were studied.

All the performance comparisons carried out in this chapter make use of result sets computed using algorithm 6. Among the parameter choices made while implementing algorithm 6 are the use of 10-second windows, 50% window overlaps and the use of 10-fold cross-validation. The reason why 10-fold cross-validation was used is explained in 3.6. However, analysis performed in section 4.3 found that a diminishing-returns relationship exists between success-rates and the length of the windows used. Similarly, analysis performed in section 4.4 found that a decreasing linear relationship exists between success-rates and logarithmic window shift while using 10-fold cross-validation. Hence, for 10-fold cross-validation, the longer the windows and the higher the window overlap; the higher the success-rates.

However, the analysis performed in this section was conducted independently of the analysis in other sections and the window lengths and window overlaps were selected independently.

No attempt is made in this chapter of finding the highest success-rate achievable by any set of parameters. Instead, the research questions answered in this chapter are strictly about the differences in the success-rates obtained by different configurations (i.e. which configuration resulted in higher success-rates than the other).

Based on the findings of section 4.3 and section 4.4, changing the window length and the window shift will almost certainly impact the success-rates obtained for any set of parameters passed to algorithm 6. However, it is hypothesised that this impact is going to be consistent across different sets of parameters and hence still lead to the same conclusions as those made in this chapter. Further research is required to verify whether this hypothesis is true or not.

In the first section of the chapter, the performance differences of the two selected feature-sets were studied. Due to the way the data was gathered, two sets of monitor setups are possible: using 3 monitors and all activities, or using 6 monitors but excluding walking and running activities. The performance of the feature-sets was studied for both setups.

The differences between success-rates obtained using the two feature-sets were found to be small but statistically significant, and larger for the 3 monitor setup than for the 6 monitor setup. The differences between success-rates obtained using the 3 monitor setup and the 6 monitor setup were found to be larger than those obtained when studying difference between the two feature-sets. The success-rates obtained using the 6 monitor setup were found to be higher than those of the 3 monitor setup.

Next, a performance comparison of success-rates obtained from accelerations, rotational velocities and orientations was performed. Again, the analysis was performed for both the 3 monitor setup and the 6 monitor setup. It was found that, of the three *sources*, accelerations are the most likely to result in the highest success-rates, and are almost certain to result in higher success-rates than rotational velocities, for either feature-set and either monitor setup. Rotational velocities, however, are most likely to result in the lowest success-rates of the three *sources* for either feature-set and either monitor setup.

A performance comparison between success-rates obtained from accelerations, rotational velocities and orientations, and those obtained from the three *sources* combined, was performed. The analysis found that, of the three *sources*, accelerations had the smallest mean difference and the lowest probability of resulting in a success-rate that was lesser than that of the combined *sources*. Orientations were had the next smallest mean difference but were either very likely to have a success-rate that is lesser than that of the combined *sources* (while using Bao and Intille's feature-set) or almost certain (while using Kwapisz et al.'s feature-set). Rotational velocities resulted in success-rates that had the greatest mean difference from that of the combined *sources* and were almost certain to have a success-rate. In addition, it was noted that the mean difference between success-rates computed from accelerations, and those computed from all three *sources*, was low, being at most 2.2%.

The implication are that, in a situations where attaining the highest success-rates is important, if all three *sources* can be obtained and processed, then combining all three is likely to result in a higher success-rate than any one individual *source*. However, if this is not possible, then accelerations should be prefered to orientations, which in turn should be preferred to rotational velocities.

Next, a comparison was performed of success-rates computed from data captured from different single body-locations. The analysis found that the wrists and the thigh are the three individual locations that result in the highest performance out of the six locations available. The differences in success-rates computed from data captured from the wrists and the thigh were found to be statistically insignificant while using Bao and Intille's feature-set. While using Kwapisz et al.'s feature-set, it was found that the dominant wrist resulted in better success-rates than the non-dominant wrist, which in turn resulted in better success-rates than the thigh.

The ranking of the ankle, chest and dominant upper arm differed depending on the feature-set used. Using Bao and Intille's feature-set, it was found that the dominant upper arm resulted in higher success-rates than both the ankle and the chest, while the ankle and the chest were similar. However, the opposite was found while using Kwapisz et al.'s feature-set: result from both the ankle and the chest were found to be higher than those of the dominant upper arm. Of the two, the ankle resulted in higher success-rates than the chest while using Kwapisz et al.'s feature-set.

This impies that if any one location on the subject's body should be selected to mount sensors to recognise Activities of Daily Living, then the wrists should be prefered to the thigh, which in turn should be preferred to the dominant upper arm, ankle or chest.

In the next section, an analysis of success-rates as a function of number of different body-locations monitored was performed. It was found that, for Bao and Intille's feature-set, while monitoring one body-location the highest success-rate was obtained if that body-location was either one of the wrists or the thigh. Using Kwapisz et al.'s feature-set, this was found to be the dominant wrist. Both feature-sets found that while monitoring two body-locations, the highest success-rate is obtained when monitoring either wrist and the thigh, and for three both wrists and the thigh.

The body-locations to be monitored for highest success-rate while monitoring four, five and six body-locations differed depending on the feature-set. For Bao and Intille's feature-set, the chest in conjunction with the wrists and thigh result in the highest success-rate for four body-locations. However, for Kwapisz et al.'s feature-set this body-location is the ankle.

It is worth noting that individually, the dominant upper arm resulted in higher success-rates than the chest while using Bao and Intille's feature-set. However, when combined with the wrists and thigh, the chest body-location resulted in slightly better (but still statistically significant) success-rates than the the dominant upper arm.

A similar observation is made with Kwapisz et al.'s feature-set where individually, the chest results in slightly higher success-rates than the dominant upper arm, but when combined with the ankle, wrists and thigh, the dominant upper arm results in slightly higher success-rates.

For five body-locations, while using Bao and Intille's feature-set it was found that there is no difference in including either the ankle or dominant upper arm to the chest, wrists and thigh. While using Kwapisz et al.'s feature-set, the dominant upper arm results in a higher success-rate when combined with the ankle, wrists and thigh for five body-locations.

The better performance of the wrists can be explained to be due to hand motions being more prevalent in Activities of Daily Living than any other body-location. The thigh performs well because of the ability to distinguish sitting from standing activities using the orientation of the monitor mounted on the thigh location.

The maximum success-rates as a function of number of body-locations were then studied. It was then hypothesised that the maximum success-rate as a function of number of body-locations forms a diminishing returns relationship. An equation was then fitted onto the data and it was observed to fit the data well. In the discussion section, this hypothesised relationship was explained to be due to less new information being made available to distinguish activities with the addition of each new monitored body-location.

The results obtained imply that, with the right placement, the first few monitors mounted on the subject's body contribute the most to the activity classification success-rate. Each additional monitor contributes lesser information than the preceeding monitor. The best locations for Activities of Daily Living are on the wrists, followed by the thigh, then either ankle, dominant upper arm or chest, depending on the feature-set used.

6 Analysis of the impact of inter-subject and inter-activity variability on activity classification accuracy

6.1 Introduction

In this chapter, analysis is performed: on the activity classification success-rates obtained by each activity to identify activities that are more easily identifiable than other activities; on the mutual confusion existing between each possible pair of activities to identify which activities are highly confusable to each other; and the success-rates obtained using 10-fold cross-validation and remove-one-subject cross-validation are analysed to understand the impact of inter-subject variation on activity classification success-rates. This chapter serves as a continuation of the analysis performed in chapter 5.

In section 6.2, success-rates of individual activities are compared to find out which activities are more easily identifiable than other activities. In addition, an analysis of mean activity success-rates with relation to standard deviations of activity success-rates is performed.

In section 6.3, mutual confusion of activities is studied to find out which activities are confused with each other. The confusability of activities is computed by using the mutual type 1 and type 2 errors between activities obtained from confusion matrices. How confusable one activity is to another is illustrated using dendrograms of activities computed from confusion matrixes.

In section 6.4, success-rates obtained using 10-fold cross-validation are compared to those of remove-one-subject cross-validation. Differences in the distributions of success-rates obtained from the two cross-validation techniques are compared. In addition, the relationship of success-rates of activities obtained from the two cross-validation techniques is evaluated.

The chapter is divided into sections. Each section contains research questions dealing with a specific area of interest. Each section begins with a discussion on the importance of studies on the area of interest; research questions in the area of interest are then possed together with reasons why we wish to attempt to answer these particular questions; a methodology of answering the research questions is given; the results of the analysis are then provided and illustrated; and finally, the conclusions and implications of the result findings are discussed. At the end of the chapter, the analysis, findings and implications of findings of the chapter are summarised.

6.2 Are some activities more easily identifiable than others?

Earlier sections in this thesis have focussed chiefly on overall activity classification success-rates. However, overall activity classification success-rates are aggregated from individual activity success-rates. This section aims to explore the relationships between success-rates obtained by individual activities in comparison to those of other activities.

To that end, the following research questions are possed:

1. **Are some activities more easily identifiable than others?**

   Knowing which activities are more easily identifiable than others can allow activity recognition researchers to focus more research on trying to distinguish the activities that are not easily identifiable.

2. **Is there a relationship between mean activity success-rates and standard deviation of activity success-rates?**

   If some activities are more easily identifiable than others, then it is likely that the activities that are easily identifiable would consistently be identified with higher success-rates, while other activities would vary hence result in a larger standard deviation in the activity success-rates.

Methodology

To answer the research questions, confusion matrices of result sets obtained in chapter 5 were analysed to extract the individual success-rates when identifying each activity. The success-rate of identifying an activity is computed based on equation 6.1.

$$\text{Success-rate}_A = \frac{count(\text{matched samples of activity } A)}{count(\text{samples of activity } A)} 100 \tag{6.1}$$

Four sets of results were computed and used to answer the research questions posed in this section. The result sets are described in table 6.1 and are similar to the result sets with similar names used in chapter 5 (refer to table 5.2 for more details).

The success-rates of activities were compared as follows: for each result-set, the success-rates of all activities were extracted. The differences between success-rates of each pair of activities was then computed. A histogram of the differences with bin size 1 was computed and a normal distribution was fitted onto the histogram using maximum likelihood estimation.

Then, using the fitted normal distribution as a probability density function, the probability of a success-rate in the first activity's result set being less than a success-rate in the second result set was computed as the cumulative distribution below zero (i.e. the integration of the fitted model for the range $(-\infty, 0]$). This probability is then used as the likelihood of the first activity having a higher activity classification

Table 6.1: Descriptions of the result sets used from chapter 5.

Result Set	Description
Bao-3	Result set obtained from feature-vectors of accelerations, rotational velocities and orientations, extracted using Bao and Intille's feature-set, and using data captured from the chest, dominant wrist, thigh.
Bao-6	Result set obtained from feature-vectors of accelerations, rotational velocities and orientations, extracted using Bao and Intille's feature-set, and using data captured from all six body-locations available, but excluding walking and running.
Kwapisz-3	Result set obtained from feature-vectors of accelerations, rotational velocities and orientations, extracted using Kwapisz et al.'s feature-set, and using data captured from the chest, dominant wrist, thigh.
Kwapisz-6	Result set obtained from feature-vectors of accelerations, rotational velocities and orientations, extracted using Kwapisz et al.'s feature-set, and using data captured from all six body-locations available, but excluding walking and running.

success-rate than the second activity given the activity classification parameters used.

Finally, the activities were ranked based on their likelihood of having a higher success-rate than other activities, given the activity classification parameters used. This process is elaborated as algorithm 8.

Results

Activity success-rates obtained

Activity success-rates from the 3 monitor setup result sets *Bao-3* and *Kwapisz-3* are summarised in the boxplot in figure 6.1. Similarly, activity success-rates from the 6 monitor setup result sets *Bao-6* and *Kwapisz-6* are summarised in the boxplot in figure 6.2.

From figure 6.1 and figure 6.2, it can be observed that activities like walking, running, walking up stairs, walking down stairs, texting on the phone, using the PC, walking (on a flat surface), watching TV and writing have high success-rates and low variance in their success-rates.

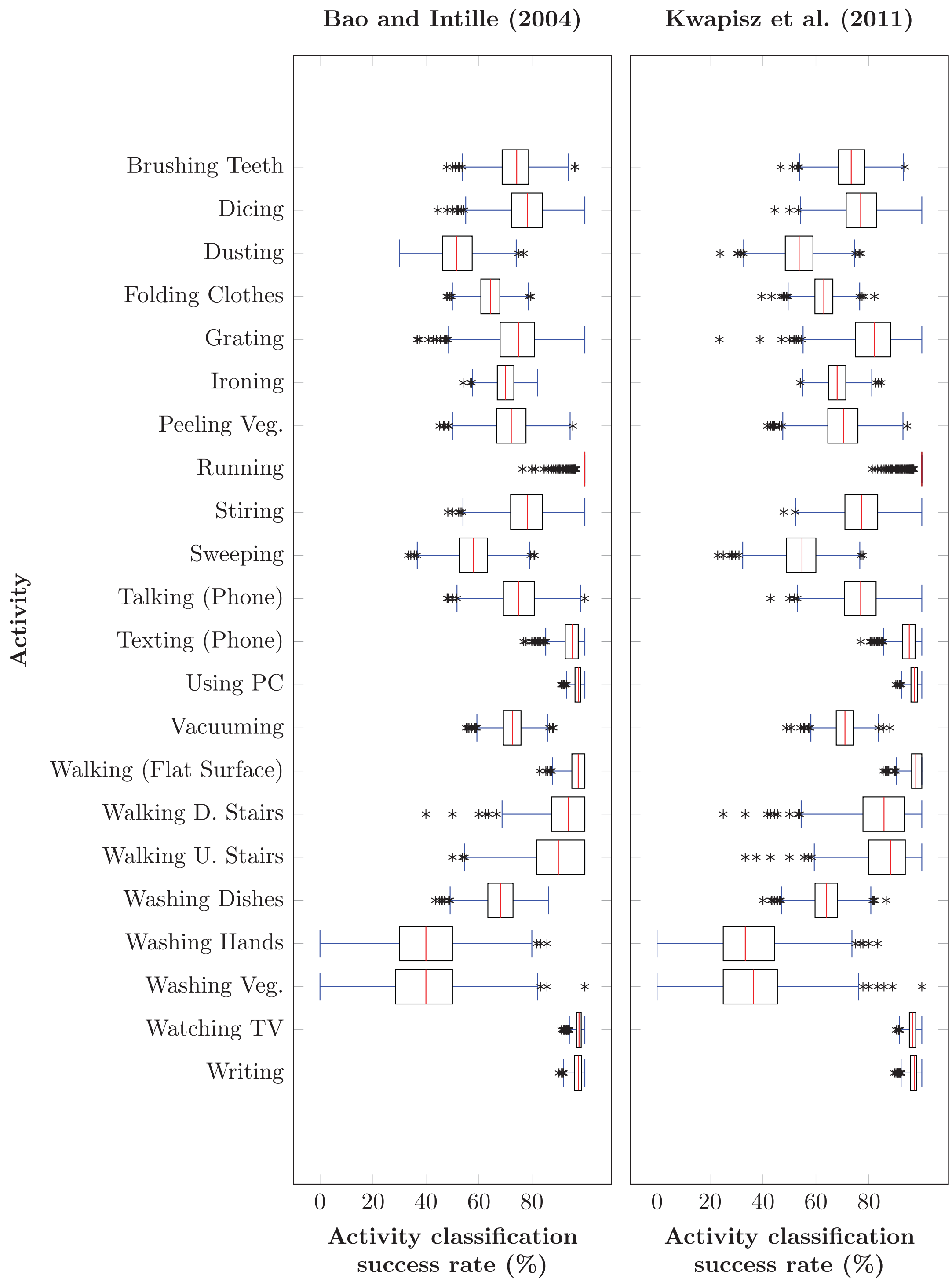


Figure 6.1: Boxplots of success-rates of each activity from *Bao-3* result set (left) and *Kwapisz-3* result set (right). The two boxplots share the y-axis on the left. Outliers are shown using asterisks.

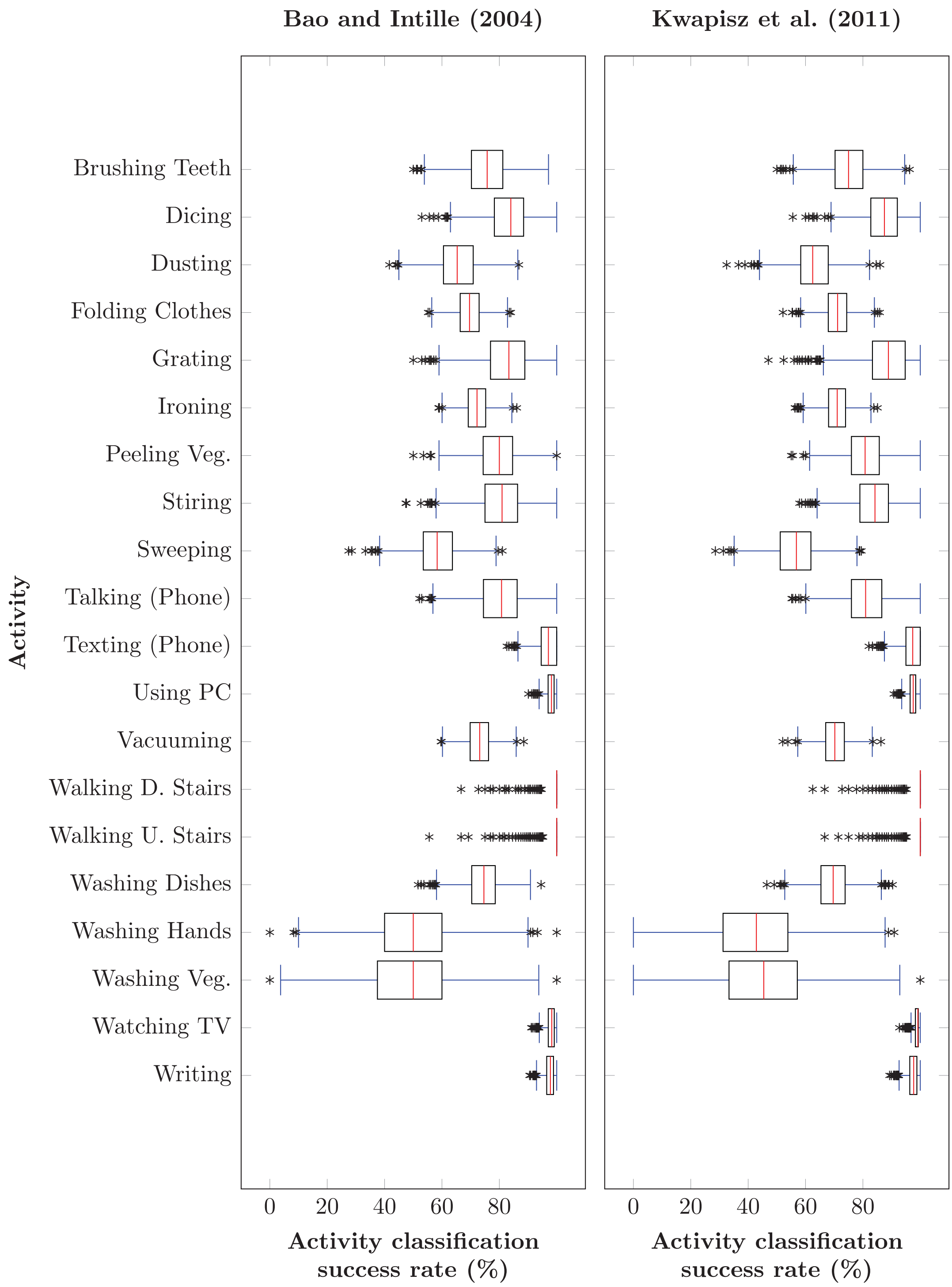


Figure 6.2: Boxplots of success-rates of each activity from *Bao-6* result set (left) and *Kwapisz-6* result set (right). The two boxplots share the y-axis on the left. Outliers are shown using asterisks.

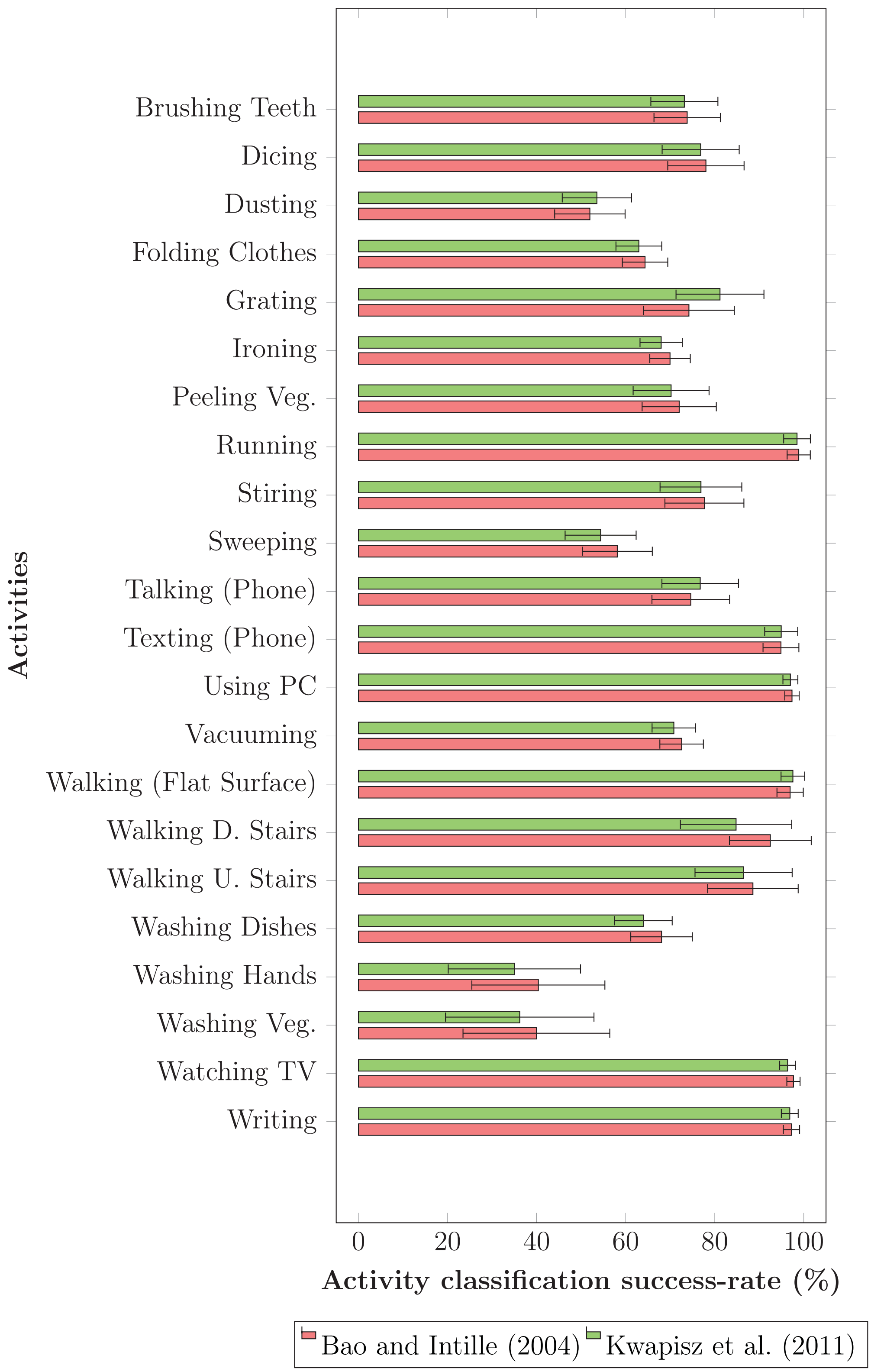


Figure 6.3: Mean activity classification success-rates (and standard deviations shown using error bars) of each activity from the *Bao-3* and *Kwapisz-3* result sets.

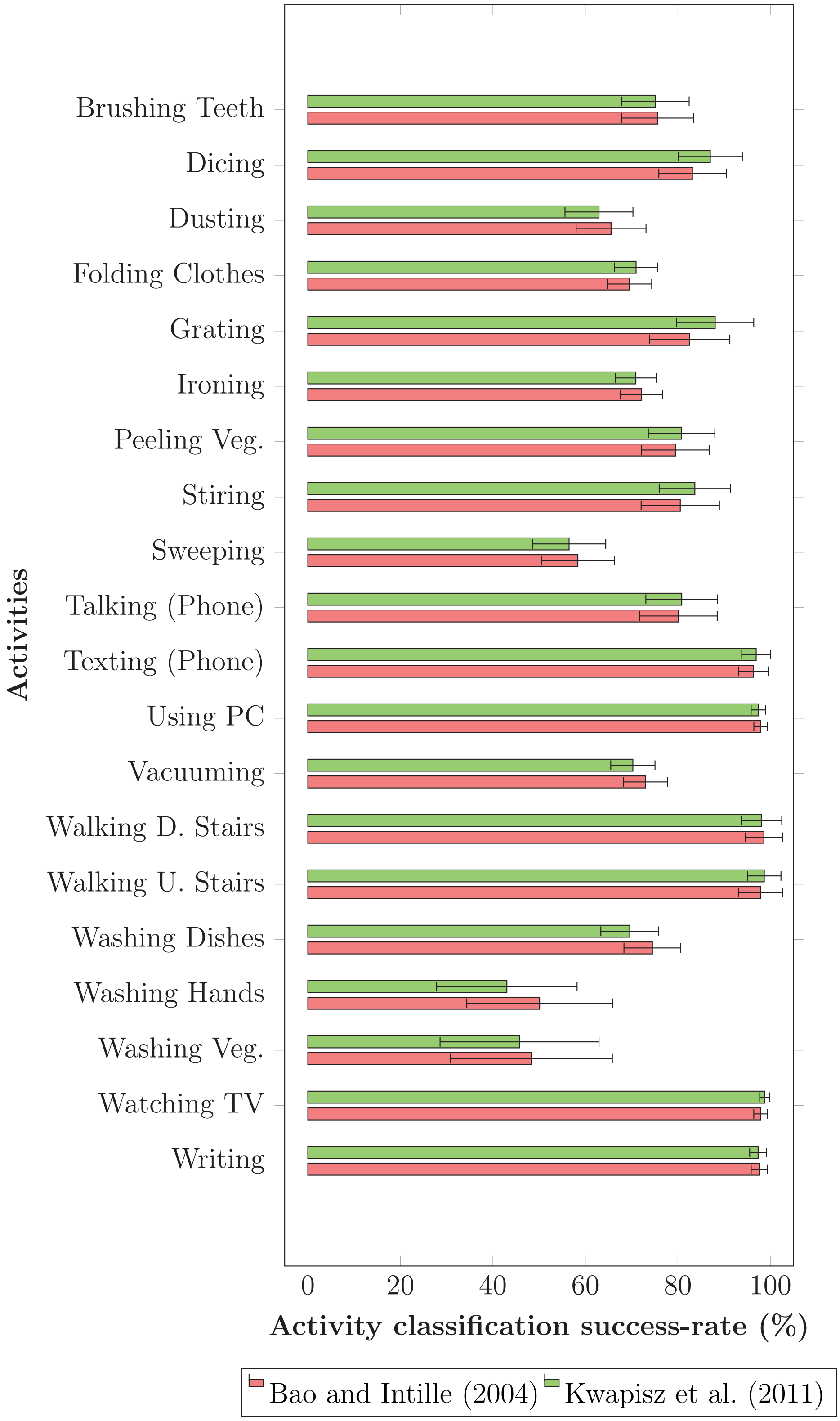


Figure 6.4: Mean activity classification success-rates (and standard deviations shown using error bars) of each activity from the *Bao-6* and *Kwapisz-6* result sets.

Algorithm 8 Rank the activities in result set $ResultSet$ based on the likelihood of one activity obtaining higher success-rates than another activity given the activity classification parameters used.

```
procedure RankActivities(ResultSet)
    Create results table T
    A ← all activities in ResultSet
    for all A1 ∈ A do
        for all A2 ∈ A, A1 ≠ A2 do
            S1 ← all success-rates of activity A1 from ResultSet
            S2 ← all success-rates of activity A2 from ResultSet
            D ← S1 − S2                          ▷ Compute difference in success-rates
            H ← Histogram of D with bin size 1
            Fit normal distribution M onto H using maximum likelihood estimate.
            Extract mean x̄ of M
            Compute the cumulative probability p at (S1 − S2) = 0 from pdf M.
            Set P(A1 > A2) ← 1 − p in table T
            Set overline(A1 − A2) ← x̄ in table T
        end for
    end for

    Sort activities A using probabilities stored in T such that P(Ai > Ai+1) >= 0.5
end procedure
```

This is easier to observe in figure 6.3 and figure 6.4, which illustrate the mean success-rates and standard deviations of the two feature-sets studied for the 3 monitor setup result sets and the 6 monitor setup result sets respectively.

From figure 6.3 and figure 6.4, it can be observed that the activities with the lowest success-rates are washing hands and washing dishes. Not only do they have the lowest success-rates, but also appear to have the highest standard deviation of success-rates.

From the result-sets, it can be observed that the success-rates of the activities obtained from different feature-sets but similar monitor setups appear very similar to each other. The correlation between success-rates of activities obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set are 0.8201 and 0.8083 using the 3 monitor setup and the 6 monitor setup respectively.

A paired two-sample two-tailed t-test with $\alpha = 0.05$ between each pair of activity result sets of the same activity and same monitor setup but computed using different feature-sets was run to test the null hypothesis that the pairs of result sets came from independent random samples from normal distributions with equal means and equal but unknown variances.

The test rejected the null hypothesis for all activities of both the 3 and the 6 monitor setups implying that the differences in the result sets are statistically significant. These results are similar to those obtained in section 5.3 that showed that the differences in the overall success-rates obtained from the two feature-sets were small but statistically significant.

Ranking activities based on likelihood of achieving higher success-rates than other activities

A paired two-sample two-tailed t-test with $\alpha = 0.05$ between each pair of activity result sets computed from the same feature-sets and the same monitor setup but of different activities was run to test the null hypothesis that the pairs of result sets came from independent random samples from normal distributions with equal means and equal but unknown variances.

The test rejected the null hypothesis for all pairs of activities implying that the differences between the success-rates obtained from the activities are statistically significant.

Hence, we can compute the differences between pairs of activities and rank them based on which activity results in higher success-rates than other activities. The rankings computed from success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set using the 3 monitor setup are given in figure D.1 and figure D.2 in appendix D. Similarly, figure D.3 and figure D.4 in appendix D give the ranking computed from success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set using the 6 monitor setup.

From figure D.1 and figure D.2 we can observe that for both feature-sets, running was the most likely to result in the highest success-rates while washing hands is the most likely to result in the lowest success-rates while using the 3 monitor setup. The activities with the lowest success-rates are observed to be in the same order for both feature-sets: washing hands, washing vegetables, dusting, sweeping, folding clothes, washing dishes, ironing, peeling vegetables, vacuuming and brushing teeth. Other activities' ranking vary depending on the feature-set used, however, some common patterns can be observed. For example, watching TV, writing, using the PC, walking on a flat surface and texting on the phone are ranked higher than all other activities except for running for both feature-sets. However, the order of the ranking within the set changes depending on the feature-set used.

Figure D.3 and figure D.4 are observed to be similar to each other. A large number of the top ranking activities have the same order for both feature-sets while using the 6 monitor setup. This order of activities is: walking up stairs, walking down stairs, watching TV, writing, using a PC, texting on the phone, grating, dicing, stiring, talking on the phone, peeling vegetables and brushing teeth. Other activities vary depending on the feature-set used. Washing vegetables and washing hands are observed to result in the lowest success-rates of all activities for both feature-sets. The next lowest are sweeping and dusting, in that order.

When the activity rankings obtained from all four result sets computed are aggregated, and activities that are likely to have higher success-rates than other activities for all result sets are obtained, the results are illustrated in figure 6.5. Figure 6.5 illustrates the rankings of all activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. The rankings generated from all four result sets (those generated from Bao and Intille's feature-set, Kwapisz et al.'s feature-set as well as 3 and 6 monitor setups) were combined. Where the rankings from one result set conflicted with those of any of the other result sets, the conflicting activities were positioned horizontally (implying that either possibility is likely).

From figure 6.5 we can observe that using the PC, watching TV and writing are likely to achieve higher success-rates than any other activity using either of the two studied feature-sets and either the 3 monitor setup or 6 monitor setup. Running is likely to perform better than using a PC, watching TV and writing using either feature-set studied, however, this was only observed from the 3 monitor setup result sets. Similarly, walking (on a flat surface) performs at the same level as using a PC, watching TV and writing, however, this was only observed from the 3 monitor setup result sets.

The activities likely to have the next highest success-rates, using either of the two feature-sets or monitor setups, are observed to be texting on the phone, walking down stairs and walking up stairs.

Washing vegetables and washing hands are observed to likely have the least success-rates of all the activities for all the result sets. Dusting and sweeping are observed to have the next highest success-rates in all the result sets followed by vacuuming, ironing, washing dishes and folding clothes.

Relationship between mean and standard deviation of activity success-rates

The mean and standard deviation of activity success-rates are observed to correlate. The correlation is higher for the 6 monitor setup than for the 3 monitor setup, having values of -0.75 and -0.70 for Bao and Intille's and Kwapisz et al.'s feature-set for the 3 monitor setup, and -0.81 for both Bao and Intille's and Kwapisz et al.'s feature-set for the 6 monitor setup.

When the mean and standard deviations of activities are plotted, the correlation is observed to result from three groups of activities for both feature-sets and for both the 3 monitor setup and the 6 monitor setup. The groups are shown in figure 6.6

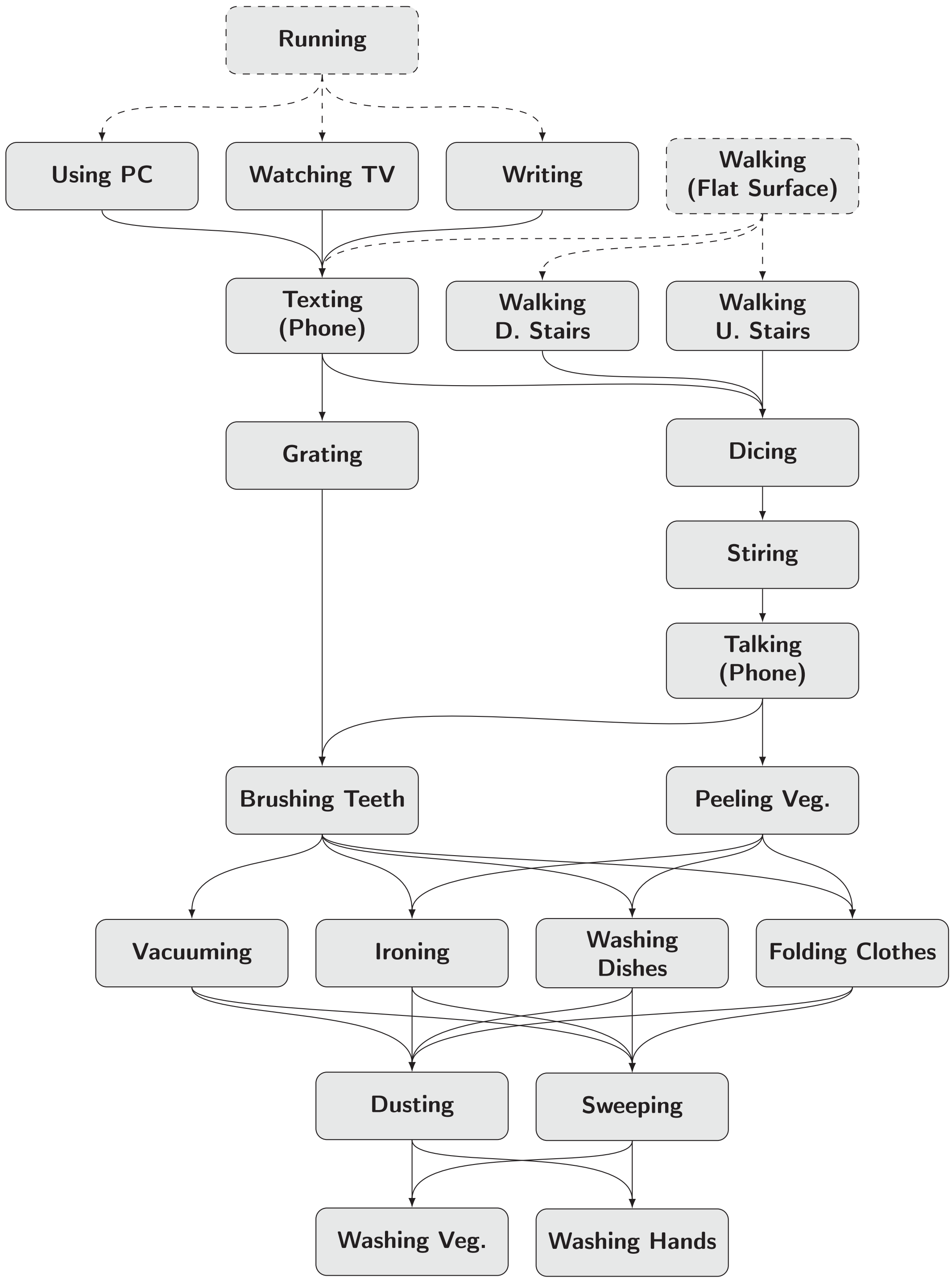


Figure 6.5: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. This illustration is an aggregation of result sets from both 3 and 6 monitor setups and both feature-sets. Only activity rankings that are valid for all result sets are shown. However, running and walking on a flat surface are also shown, even though their ranking is only based on the 3 monitor setup result sets.

and figure 6.7 for the 3 monitor setup and the 6 monitor setup respectively and can be summarised as:

1. A group with high success-rates and low variance: This group includes walking up stairs, walking down stairs, watching TV, writing, using the PC and texting on the phone. It also includes walking on a flat surface and running for the 3 monitor setup.
2. A group that has low success-rates and high variance: This group includes washing hands and washing vegetables.
3. A group that has medium success-rates and medium variance: This group includes all the other activities: brushing teeth, dicing, dusting, folding clothes, grating, ironing, peeling vegetables, stiring, sweeping, talking on the phone, vacuuming and washing dishes.

Discussion

In this section, success-rates of activities were compared to success-rates of other activities.

First, success-rates of activities were computed using both Bao and Intille's feature-set and Kwapisz et al's feature-set, and for the 3 monitor setup and the 6 monitor setup.

A *t*-test was then performed between activity success-rates obtained from the two feature-sets for both the 3 monitor setup and the 6 monitor setup. The test rejected the null hypothesis that the activity success-rates obtained from the two feature-sets came from independent random samples from normal distributions of equal mean and equal but unknown variances. Hence, the differences in the success-rates obtained from the two feature-sets are statistically significant, although they are observed to be small. These results reflect the results obtained in section 5.3 that showed that the differences in the overall success-rates obtained from the two

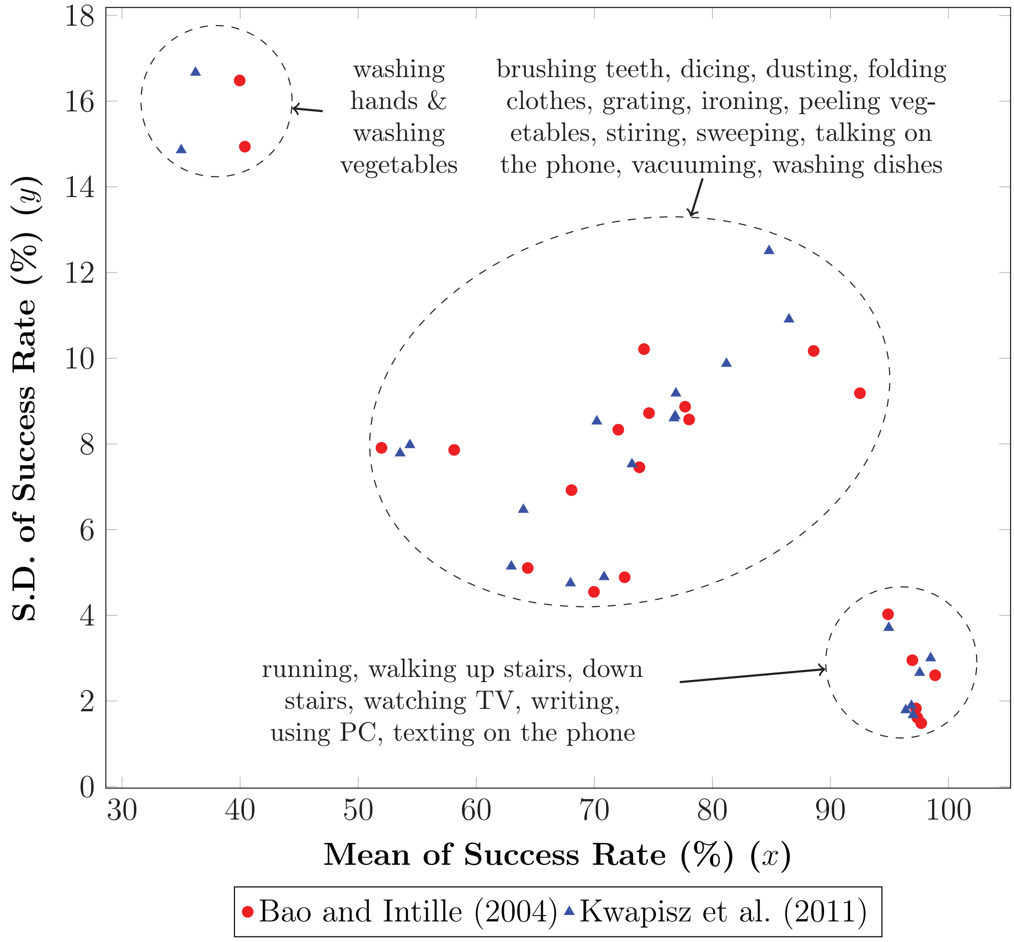


Figure 6.6: Scatter plot of mean and standard deviations of success-rates obtained by activities using Bao and Intille's feature-set and Kwapisz et al's feature-set for the 3 monitor setup. The correlation coefficient between x and y success-rates obtained using Bao and Intille's feature-set is -0.7482, while the correlation coefficient between x and y success-rates obtained using Kwapisz et al.'s feature-set is -0.6992.

feature-sets were small but statistically significant.

Similar t-tests were then performed between result sets of each possible pair of activities of the same feature-set and same monitor setup. The t-tests rejected the null hypothesis that the activity success-rates came from independent random samples of normal distributions of equal mean and equal but unknown variance. Hence, the differences in the success-rates between each possible pair of activities of the same feature-set and the same monitor setup, were found to be statistically significant.

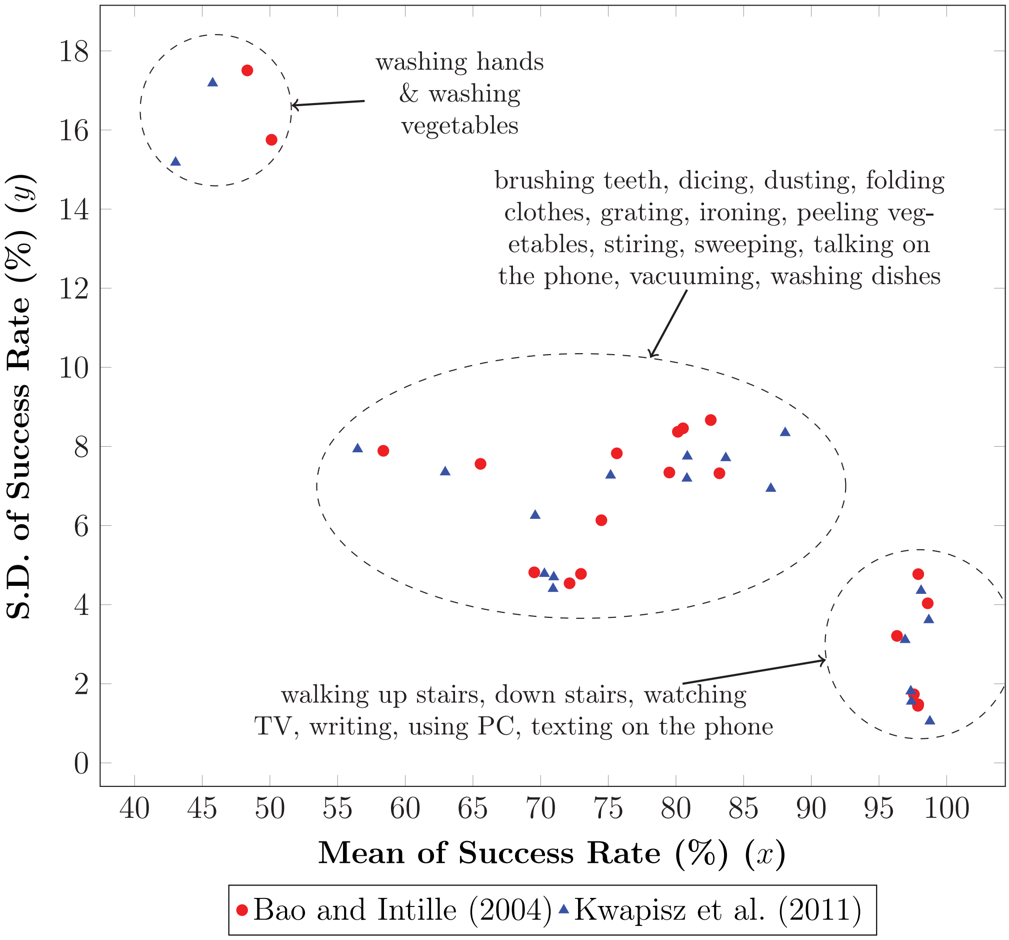


Figure 6.7: Scatter plot of mean and standard deviations of success-rates obtained by activities using Bao and Intille's feature-set and Kwapisz et al's feature-set for the 6 monitor setup. The correlation coefficient between x and y success-rates obtained using Bao and Intille's feature-set is -0.8130, while the correlation coefficient between x and y success-rates obtained using Kwapisz et al.'s feature-set is -0.8065.

The differences between the activity success-rates were then computed. A histogram of the differences was then computed and a normal distribution fitted onto the histogram using maximum likelihood estimation. The fitted normal distribution was then used as a probability density function and the probability of the first activity's result set being less than the second result set was computed as the cumulative distribution below zero (i.e. the integration of the fitted model for the range $(-\infty, 0])$. Using the probabilities, the activities were ranked such that activities that were more likely of obtaining higher success-rates than other activities were ranked higher than the other activities. This was performed for both feature-sets and for both the 3 and 6 monitor setup.

From the 3 monitor setup rankings, it was noted that running was most likely to have the highest success-rate while washing hands was most likely to have the lowest success-rate. It was also noted that the lowest performing activities while using the 3 monitor setup had the same order for both feature-sets. The order (from the lowest) is: washing hands, washing vegetables, dusting, sweeping, folding clothes, washing dishes, ironing, peeling vegetables, vacuuming and brushing teeth. The ranking of other activities changed depending on the feature-set used.

From the 6 monitor setup rankings, it was noted that the order of the activities with the highest success-rates were similar between the two feature-sets studied. The order (from the highest) was: walking up stairs, walking down stairs, watching TV, writing, using the PC, texting on the phone, grating, dicing, stirring, talking on the phone, peeling vegetables, and brushing teeth. The rankings of other activities changed depending on the feature-set used.

When the rankings obtained from the four result sets were combined, it was observed that using the PC, watching TV and writing were likely to have higher success-rates than any other activity. Running is likely to be higher than the three, but this observation was only made for the 3 monitor setup. Similarly, walking on a flat surface is likely to have as high success-rates as the three, but this observation was only made for the 3 monitor setup. The activities likely to have the next highest success-rates are observed to be texting on the phone, walking down stairs and walking up stairs.

Washing vegetables and washing hands are likely to have the lowest success-rates; followed by dusting and sweeping; and vacuuming, ironing, washing dishes and folding clothes. The full ranking is illustrated as figure 6.5.

Next, it was observed that a correlation existed between the mean and standard deviation of success-rates obtained from activities. The correlation coefficient of success-rates obtained using Bao and Intille's feature-set was -0.7482 and that of Kwapisz et al.'s feature-set was -0.6992 while using the 3 monitor setup. The correlation coefficient of success-rates obtained using Bao and Intille's feature-set was -0.8130 and that of Kwapisz et al.'s feature-set was -0.8065 while using the 6 monitor setup. Hence, the correlation was observed to be stronger for the 6 monitor setup than for the 3 monitor setup.

When the activity mean success-rates were plotted against the standard deviations of the success-rates, it was observed that the correlation is a result of the existence of three distinct groups of activities: those with high success-rates and low variance, those with low success-rates and high variance, and those with medium success-rates and medium variance. The three groups were observed while using both feature-sets and while using both monitor setups.

The group of activities with the low success-rates and high variance is observed to consist of washing hands and washing vegetables. The success-rates of these activities largely vary from one testing fold to another, resulting in higher success-rates in some folds and lower success-rates in other folds. This is possibly because some sections of these activities are much easier to identify than other sections using the feature-sets studied. Hence, if the testing fold contains windows from the easier-to-identify sections, the testing fold results in higher success-rates, while if the testing fold has no windows of the easier-to-identify sections, the testing fold results in lower success-rates.

The group of activities with the high success-rates and low variance are observed to consist of walking up stairs, walking down stairs, watching TV, writing, using the PC and texting on the phone. This group also includes walking on a flat surface and running for the 3 monitor setup. These activities are consistently identified accurately. Hence, all or most windows of these activities can be identified accurately. This is possibly because these activities have particular characteristics that the feature-sets are consistently capturing to identify the activities.

For example, running is known to have high accelerations due to high impact with the ground. This distinguishing characteristic results in much higher amplitudes in some of the acceleration FFT bins than others and is captured as a higher information entropy in the frequency domain while using Bao and Intille's feature-set. This distinguishing characteristic of running also results in a higher standard deviation of the acceleration signal which is captured by the standard deviation of the signal or the standard deviation of the magnitude of the signal by Kwapisz et al.'s feature-set. Similarly distinct characteristics could exist for the other activities too.

The results of this section imply that activities like walking on a flat surface, walking up stairs, walking down stairs, running, watching TV, writing, using a PC and texting on the phone are easier to identify and hence activity recognition systems that only identify these activities are likely to have high success-rates. However, activities like washing hands, washing dishes, dusting and sweeping are more difficult to identify.

6.3 Activity classification specificity

The motions that define some activities are more alike than others. For example, we expect the hand movements involved in washing vegetables and washing dishes to be more similar than those between washing vegetables and walking up stairs.

However, when algorithms are compared (as they usually are in papers) the comparison is based on the average of the success rate of recognising each activity. However, this does not take into account, as explained in Consolvo et al. (2008) and Taylor et al. (2011), that this approach is different from the users' perception of accuracy. To achieve a high level of user acceptance, only activities that can be determined by both a high level of sensitivity and a high level of specificity should be reported (Taylor et al., 2011).

A balance needs to be struck on how specific an activity recognition system is when it is reporting activities identified. An activity recognition system that is very specific in reporting activities identified, at the risk of being inaccurate, results in a poor user perception of the system. On the other hand, a system that is vague in it's reporting of the user's activities is unlikely to be useful to the user.

Hence, if a system can not determine accurately enough whether a user is washing vegetables or washing dishes, it can instead report a combined class (perhaps *washing at the sink*). This combined class is more likely to be correct than either of the two more specific options. However, at the most extreme, it might not be useful to the user. For example, *physical exercise* is not useful for a user who expects to keep track of which exercise he or she performed and hence how many calories were burnt over the exercise duration.

The ideal activity recognition system would be able to perfectly recognise any number of activities, however, the reality is that this is a difficult task to accomplish given the variety of activities undertaken by end-users. This makes the criteria used to pick an activity recognition algorithm important.

Hence when selecting an activity recognition algorithm based on it's success rate, an activity recognition algorithm that confuses washing vegetables with washing dishes is better than one that confuses washing vegetables with walking up stairs. This is because washing vegetables and washing dishes can naturally be grouped into a higher level activity that the user can accept while washing vegetables and walking up stairs can not.

To that end, the research question posed in this section is: which activities are confused with one another using the feature-sets studied? Hence, in this section, the results from the two selected feature-sets are analysed to find out how much confusion exists between different activities.

To aid in the analysis, the data is presented in the form of an *activity dendrogram*. A dendrogram is a tree diagram that is often used to illustrate hierarchical clustering. We propose an *activity dendrogram* as a dendrogram that illustrates confusability between activities with relation to a particular activity-classification algorithm. We base the confusability of activities on the notion that highly confusable activities result in higher mutual type 1 and type 2 errors in classification.

Methodology

Confusion matrixes computed using algorithm 6 were obtained so as to compare mutual confusability of activities while using result sets obtained using both Bao and Intille's feature-set and Kwapisz et al.' feature-set for:

1. The 3 monitor setup compared to the 6 monitor setup.
2. Acceleration, rotational velocities and orientations.

Hence, 16 sets of results were computed using algorithm 6. The result sets are similar to those used in section 5.4. Refer to table 5.4 for details of the parameters passed to algorithm 6 to compute the result sets.

To create activity dendrograms using the confusion matrices obtained from the result sets, agglomerative hierarchical clustering was used. Agglomerative hierarchical clustering is a form of bottom-up clustering where at each step, the two subgroups with the shortest distance are merged (or linked) to form a higher-level group resulting into a hierarchy of clusters. In our case, we wish to merge the two groups of activities with the highest mutual type 1 and type 2 error rates. By doing so, we create a hierarchy that at the bottom has each activity in its own group. At each step, the most similar activity groups (i.e. those with the highest mutual error rates) are merged upwards hence reducing the overall error rates and increasing the overall classification success-rates.

Algorithm 9 elaborates on how the activity dendrograms were created.

Algorithm 9 Algorithm that creates an activity dendrogram from a confusion matrix.

```
procedure CreateActivityDendrogram(M)
    node_map ← new Map                          ▷ Map to store dendrogram nodes in.
    for all a ∈ activities(M) do                ▷ Add all activities to map.
        node_map.add(a, new Node(a))
    end for

    while node_map.size > 1 do                  ▷ Loop till only one activity is left.
        activity1 ← NULL                        ▷ Data types to store the pair of
        activity2 ← NULL                        ▷ activities with the largest error.
        largest_error ← −∞

        for all a ∈ activities(M) do                ▷ Find the pair of activities with
            for all b ∈ activities(M), a ≠ b do         ▷ the largest error.
                e ← error(a, b) + error(b, a)       ▷ Add type 1 and type 2 errors.
                if e > largest_error then
                    largest_error ← e
                    activity1 ← a
                    activity2 ← b
                end if
            end for
        end for

                                                ▷ Merge activities to create parent activity.
        parent_activity ← activity1 + activity2
                                                ▷ Update activities in matrix.
        M ← merge(M, activity1, activity2, parent_activity)
                                                ▷ Compute success rate from new matrix.
        success_rate ← calc_success_rate(M)
                                                ▷ Create new dendrogram node for parent.
        parent_node ← new Node(parent_activity)
        node_map.add(a, parent_node)                ▷ Add parent node to node map.
                                                ▷ Set parent of activity nodes.
        node_map[activity1].parent ← parent_node
        node_map[activity2].parent ← parent_node

        node_map.remove(activity1)              ▷ Remove child activities
        node_map.remove(activity2)              ▷ from the node map.
    end while

    return node_map.first                       ▷ Return the root of the dendrogram.
end procedure
```

Results

The activity dendrograms computed from result sets in this section are given in appendix E.

Figure 6.8 and figure 6.9 the generated dendrograms based on mutual error rates between each possible pair of activities obtained from results obtained from Bao and Intille's feature-set and Kwapisz et al.'s feature-set using the 3 monitor setup (figure 6.8) and 6 monitor setup (figure 6.9) and all three *sources*.

It should be noted that the y-axis of the dendrograms present the success-rate obtained when the data of subclusters are merged and the resulting data reclassified.

Figure 6.8 and figure 6.9 show activity dendrograms constructed from confusion matrixes obtained from classifying feature-vectors extracted from all three *sources*. Figure 6.8 was constructed from result sets *Bao-3* and *Kwapisz-3* while figure 6.9 was constructed from result sets *Bao-6* and *Kwapisz-6*.

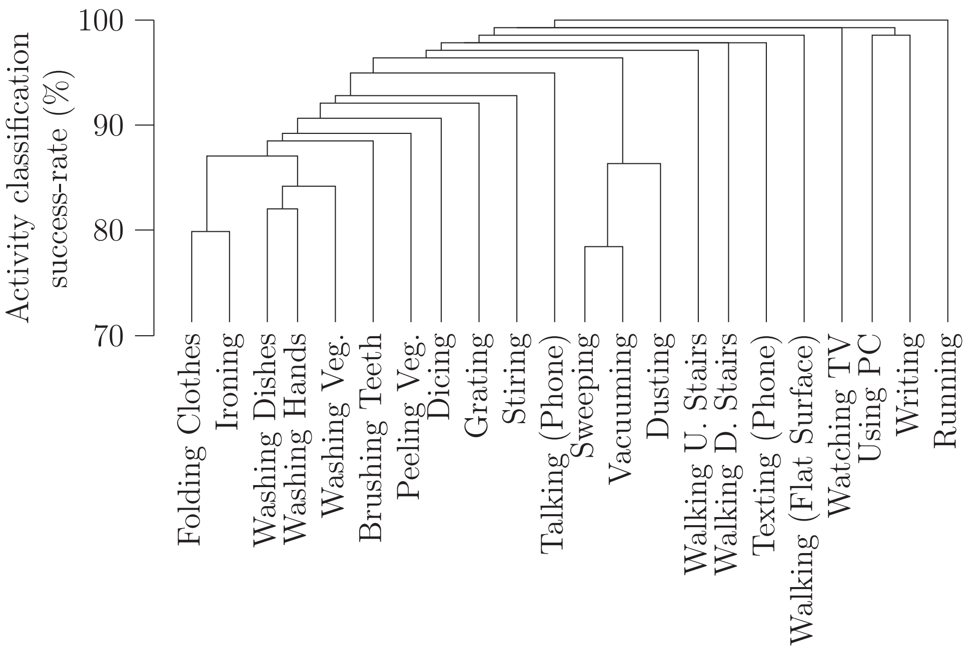
Figure 6.8 and figure 6.9 are very similar. Many of the clusters visible in figure 6.8 can also be seen in figure 6.9. The most visible difference between the two figures is that success-rates in figure 6.8 are lower than those of figure 6.9.

From the two figures, it can be observed that the greatest mutual confusion exists between sweeping and vacuuming for both feature-sets and both the 3 and 6 monitor setups. The pair of activities with the next highest mutual confusion depends on the feature-set used. The pair is folding clothes and ironing clothes for Bao and Intille's feature-set for both monitor setups and washing hands and washing dishes for Kwapisz et al.'s feature-set for both monitor setups.

The common three groups that result in the highest mutual confusion for both feature-sets and both 3 monitor setups and 6 monitor setups are: sweeping, vacuuming and dusting; folding clothes and ironing clothes; and washing dishes, washing hands and washing vegetables.

Figure E.4, figure E.5 and figure E.6 show activities clustered based on mu-

Bao and Intille (2004)



Kwapisz et al. (2011)

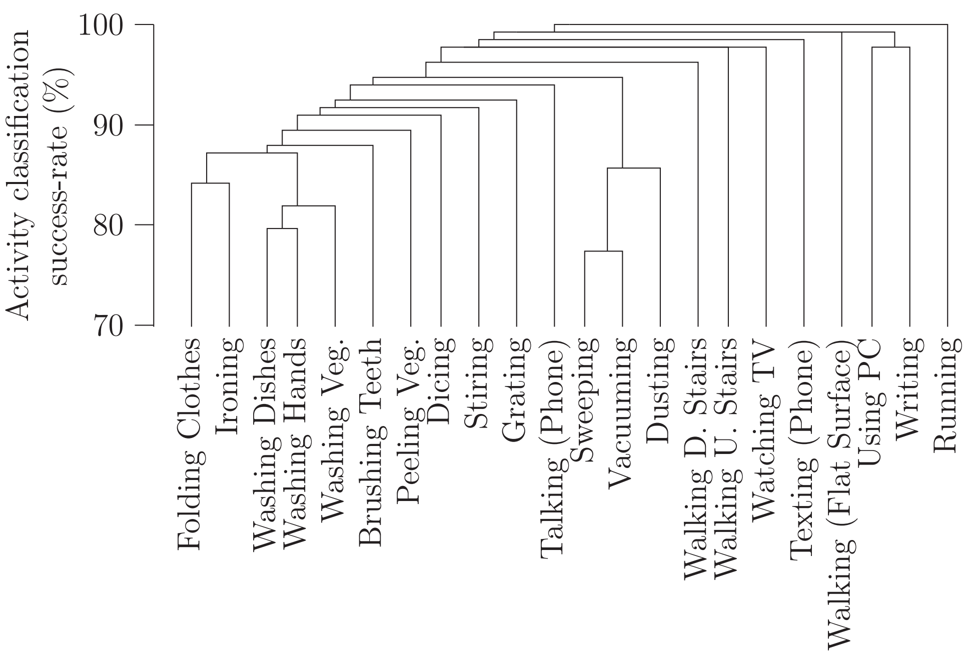


Figure 6.8: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-3* (upper) and *Kwapisz-3* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.

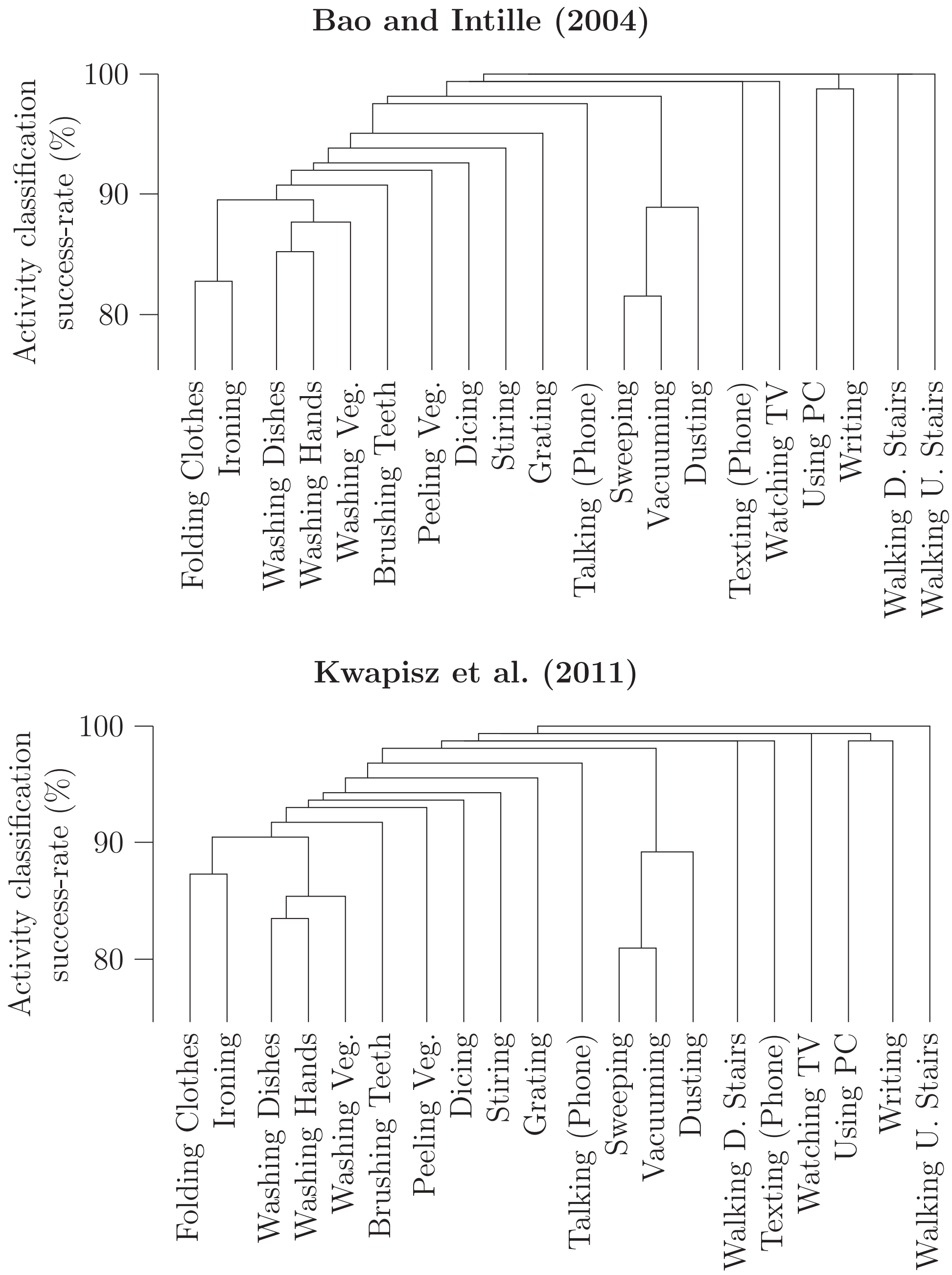


Figure 6.9: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-6* (upper) and *Kwapisz-6* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.

tual confusability computed from result sets obtained from accelerations, rotational velocities and orientations, respectively, using the 6 monitor setup. Similarly, figure E.1, figure E.2 and figure E.3 show activities clustered based on mutual confusability computed from result sets obtained from accelerations, rotational velocities and orientations, respectively, using the 3 monitor setup.

With a few exceptions, the figures are mostly similar to each other and to the combined accelerations, rotation-velocities and orientation figure 6.9. Similar to figure 6.8 and figure 6.9, the greatest mutual confusion exists between sweeping and vacuuming for all result sets obtained from individual *sources*. Similarly, the next pair of activities is either observed to be folding and ironing or washing dishes and washing hands depending on the result set. All result sets from individual *sources* except for the result set computed from Kwapisz et al.'s feature-set using orientations, have folding and ironing as the pair of activities with the next highest mutual confusion.

Another commonality between activity dendrograms of results sets computed from individual *sources* to those of all three *sources* combined, is the presence of the three groups of activities that result in high confusion: sweeping, vacuuming and dusting; folding clothes and ironing clothes; and washing dishes, washing hands and washing vegetables.

One of the notable differences between the figures computed from rotational velocities from those computed from accelerations and orientations is the decreased capability to distinguish between using a PC, watching TV and writing while using rotational velocities. Additionally, the capability to distinguish between talking on the phone and texting on the phone is observed to be lesser while using Kwapisz et al.'s feature-set than while using Bao and Intille's feature-set for rotational velocities.

Discussion

In this section, the research question posed was which activities are confused with one another using the feature-sets studied. To answer the research question, activity classification confusion matrixes were used to find which activities had the highest mutual error rates. The activities with the highest mutual error rates were then merged and the process was repeated for the next pair of activities with the highest mutual error rates. The result of this process were expressed as dendrograms showing activities that result in the highest mutual confusion using the feature-set being getting paired lower(on the figure) than other activities. This process was repeated for both the 3 and 6 monitor setups using all three *sources* combined and for each individual *source* using the 6 monitor setup.

From the activity dendrograms generated, it was observed that, for result sets from both the 3 and the 6 monitor setups, both the feature-sets, and using either all three *sources* combined or individually, the greatest mutual confusion is found between sweeping and vacuuming.

This is because sweeping and vacuuming have similar motions. It is likely that of all the activities, sweeping and vacuuming have the most similar motions to each other. To improve their identification, it is possible that additional sensors that identify the presence of the vacuum cleaner would assist in discriminating vacuuming from sweeping. Such sensors could be RFID readers, or the use of a microphone to sense ambient noise levels. While both approaches require the additional hardware, the second approach is far easier to deploy because all mobile phones have microphones and many people carry mobile phones. Hence it would simply require the system developer to make use of the microphone on the user's phone to sense the ambient noise levels to determine whether the user is sweeping or vacuuming. However, this method would not work if the user was sweeping in the same room as someone else who is vacuuming and would likely not work well in a noisy environment.

The next pair of activities with the highest mutual confusion is observed to be either folding clothes and ironing or washing hands and washing dishes depending on the result set. Between ironing and folding clothes, ironing includes in it folding motions and clothes flipping motions. These motions are prevalent in folding but are also part of ironing and it is not possible to have ironing without them. Hence, to improve the identification of ironing from folding clothes, it is necessary to focus on the identification of the motions that are found in ironing but not in folding. Such motions include gliding the iron over the clothes. If the activity is found to be either ironing or folding by using feature-sets as those used in the analysis in this section, an additional step could be to detect the presence of these motions. Their detection implies the likelihood of the activity being ironing and not folding, while the absence implies the likelihood of the activity being folding and not ironing. Similarly, between washing hands and washing dishes, it possible that some differences in the motions of the two activities exist. For example, the scrubbing action would only be found when the user is washing dishes but not while washing hands. The isolation of these motions that distinguish one activity from another is an ideal application of time-series shapelets (Ye & Keogh, 2009). Time-series shapelets refer to sections of time-series that are more likely to be found in one class of time-series than in another and hence can be used to distinguish such similar activities.

Common groups of activities that result in high confusion were also observed in all result sets studied. These groups are: sweeping, vacuuming and dusting; folding clothes and ironing clothes; and washing dishes, washing hands and washing vegetables. Interestingly, these groups of activities can be grouped into less specific groups: house-hold cleaning; post-laundry activities; and washing at the sink. Using these groups instead of the more specific activities would increase the overall success-rates of the activity recognition system. From figure 6.8 and figure 6.9, we can observe that merging folding, ironing clothes, washing dishes, washing hands and washing vegetables results in success-rates in excess of 80% for the 3 monitor setup and close to 90% for the 6 monitor setup for both feature-sets studied.

Within results sets computed from individual *sources*, a decreased capability to distinguish between using PC, writing and watching TV is observed in result sets computed from rotational velocities with either feature-sets. This is likely because these activities are mostly stationary (in terms of moving limbs) and hence highly depend on the feature-set's ability to distinguish the orientation of the body-locations. While orientations by definition are the orientations of the monitor mounted on the body-locations, and accelerations encode in them the orientation of the monitor mounted on the body-location through the gravity vector (see Mizell (2003)), rotational velocities don't encode that information. Hence while the accelerations and orientations captured during these activities are likely to differ from activity to activity, the rotational velocities would have mean values close to zero.

6.4 Remove-one-subject cross-validation compared to 10-Fold cross-validation?

As explained in section 3.6, many cross-validation techniques exist in activity recognition literature: N-fold and remove-one-subject. In the analyses performed in this thesis, 10-fold cross-validation was prefered due to remove-one-subject cross-validation because it was observed that results obtained from 10-fold cross-validation had less variability, higher statistical significance and therefore clearer and easier to derive conclusions from than those of remove-one-subject cross-validation.

In this section, result sets from the two cross-validation strategies are compared to evaluate the differences in overall activity classification success-rates and individual activity success-rates.

The research questions posed in this section are:

1. **How different are activity classification success-rates obtained from 10-fold cross-validation from those of remove-one-subject cross-validation?**
   Understanding the differences in activity classification success-rates to be expected between 10-fold cross-validation from those of remove-one-subject cross-validation is important because it allows activity recognition researchers a deeper understanding when comparing results generated from the two cross-validation methods.
2. **Are activity classification success-rates obtained from 10-fold cross-validation correlated to those obtained from remove-one-subject cross-validation?**
   Understanding whether or not activity classification success-rates obtained from 10-fold cross-validation correlate with those obtained from remove-one-subject cross-validation is important because, if the two sets of results correlate it could imply a relationship between the results sets. More specificly, it could imply a linear relationship which would allow activity recognition researchers a better understanding of how results obtained from 10-fold cross-validation map to those obtained from remove-one-subject cross-validation and vice versa.

Methodology

In other sections, analysis was conducted focussing on factors that impact activity classification success-rates by comparing sets of data that were produced from identical raw data and in a method that was identical except for that factor.

For example, while analysing the impact of window lengths, result sets were obtained that had been computed via methods that were identical in all aspects except for having different window lengths. Hence, success-rates from fold 1 of window length 5 seconds could be compared to those of fold 1 of window length 10 seconds. This paired analysis or testing is made possible because the same raw data that went into computing fold 1 of window length 5 seconds also went into computing fold 1 of window length 10 seconds.

However, in the case of comparing the results of remove-one-subject cross-validation to 10-fold cross-validation, not only do different sets of raw data go into computing the different testing folds, but also the cross-validation strategies result in a different number of folds. Therefore, care is taken not to perform any paired analysis on the success-rates obtained from each fold. However, comparison is made using the distributions of the overall success-rates and the success-rates of activities to compare the success-rates obtained for the activities tested using the two cross-validation strategies.

Hence, the analysis of the success-rates obtained using the two cross-validation strategies is limited to the characteristics of the distributions of the success-rates obtained, and the correlation of activity success-rates, obtained from the two cross-validation strategies.

To answer the research questions posed, algorithm 6 was altered so as to allow either 10-fold cross-validation or remove-one-subject cross-validation. Four result sets were then computed using parameters described in table 6.2.

Results

Figure 6.10 shows a boxplot of the result sets *Bao-NFold*, *Bao-Rem1Sub*, *Kwapisz-NFold* and *Kwapisz-Rem1Sub*. From the boxplot, it can be observed that the success-rates obtained using 10-fold cross-validation have a higher median and have a smaller interquartile range than those obtained using remove-one-subject for both Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

Figure 6.11 elaborates further. Figure 6.11 shows the mean, median, mode, range, inter-quartile range and standard deviation. All the measures of central tendency (mean, median and mode) computed from success-rates obtained using

Table 6.2: Result sets computed using algorithm 6 so as to analyse the differences in success-rates between those obtained using 10-fold cross-validation and those obtained using remove-one-subject cross-validation.

Result Set	**Parameters**				**Cross Validation**
	FeatureSet	***Activities***	***Sources***	***Monitors***	
Bao-NFold	Bao and Intille (2004)	All except walking and running	All 3 *sources*	All 6 available monitors	10-fold
Kwapisz-NFold	Kwapisz et al. (2011)	All except walking and running	All 3 *sources*	All 6 available monitors	10-fold
Bao-Rem1Sub	Bao and Intille (2004)	All except walking and running	All 3 *sources*	All 6 available monitors	Remove-One-Subject
Kwapisz-Rem1Sub	Kwapisz et al. (2011)	All except walking and running	All 3 *sources*	All 6 available monitors	Remove-One-Subject

the 10-fold cross-validation are higher than those computed from success-rates obtained using the remove-one-subject cross-validation for both feature-sets studied. However, all the measures of dispersion (range, interquartile-range and standard deviations) computed from success-rates obtained using the 10-fold cross-validation are lower than those computed from success-rates obtained using the remove-one-subject cross-validation for both feature-sets studied.

Figure 6.10 shows a boxplot of the result sets *Bao-NFold*, *Bao-Rem1Sub*, *Kwapisz-NFold* and *Kwapisz-Rem1Sub*. From the boxplot, it can be observed that the success-rates obtained using 10-fold cross-validation have a higher median and have a smaller interquartile range than those obtained using remove-one-subject for both Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

Chi-square goodness of fit tests with $\alpha = 0.05$ testing the null hypothesis that the success-rates in the result sets *Bao-NFold*, *Kwapisz-NFold*, *Bao-Rem1Sub* and *Kwapisz-Rem1Sub* fit normal distributions with a means and variances estimated from the result sets found that the result sets *Bao-NFold* (df=5, Chi

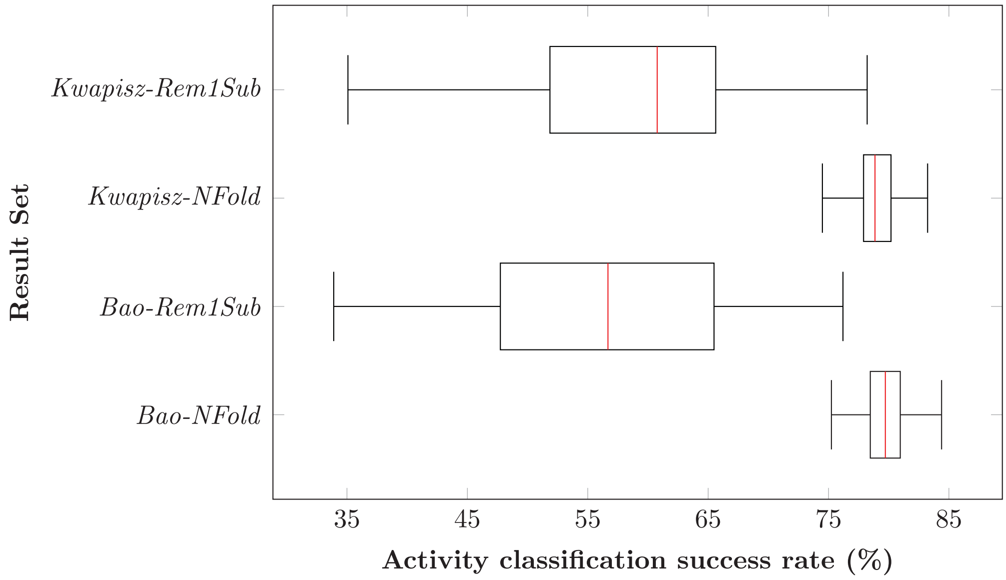


Figure 6.10: Boxplot of overall success-rates obtained from Bao and Intille's feature-set and Kwapisz et al.'s feature-set using 10-fold cross-validation (*Bao-NFold* and *Kwapisz-NFold* respectively) and remove-one-subject cross-validation (*Bao-Rem1Sub* and *Kwapisz-Rem1Sub* respectively).

Square=5.0545) and *Kwapisz-NFold* (df=5, Chi Square=9.5112) fit normal distributions but *Bao-Rem1Sub* (df=7, Chi Square=33.4837) and *Kwapisz-Rem1Sub* (df=7, Chi Square=52.2770) do not.

Figure 6.12 shows a boxplot of individual activity success-rates obtained from result sets *Bao-NFold* and *Bao-Rem1Sub*. Similarly, figure 6.13 shows a boxplot of the activity success-rates obtained from result sets *Kwapisz-NFold* and *Kwapisz-Rem1Sub*. From the figures, it can be observed that the same trend observed in figure 6.10 occurs at the individual activity level. From the boxplots, it can be observed that the success-rates obtained using 10-fold cross-validation have a higher median and have a smaller interquartile range than those obtained using remove-one-subject for both Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

Figure 6.14 shows a scatter plot of the medians of individual activity success-rates obtained using remove-one-subject cross-validation plotted against the medians of individual activity success-rates obtained using 10-fold cross-validation. The two sets of median activity success-rates are observed to have a strong correla-

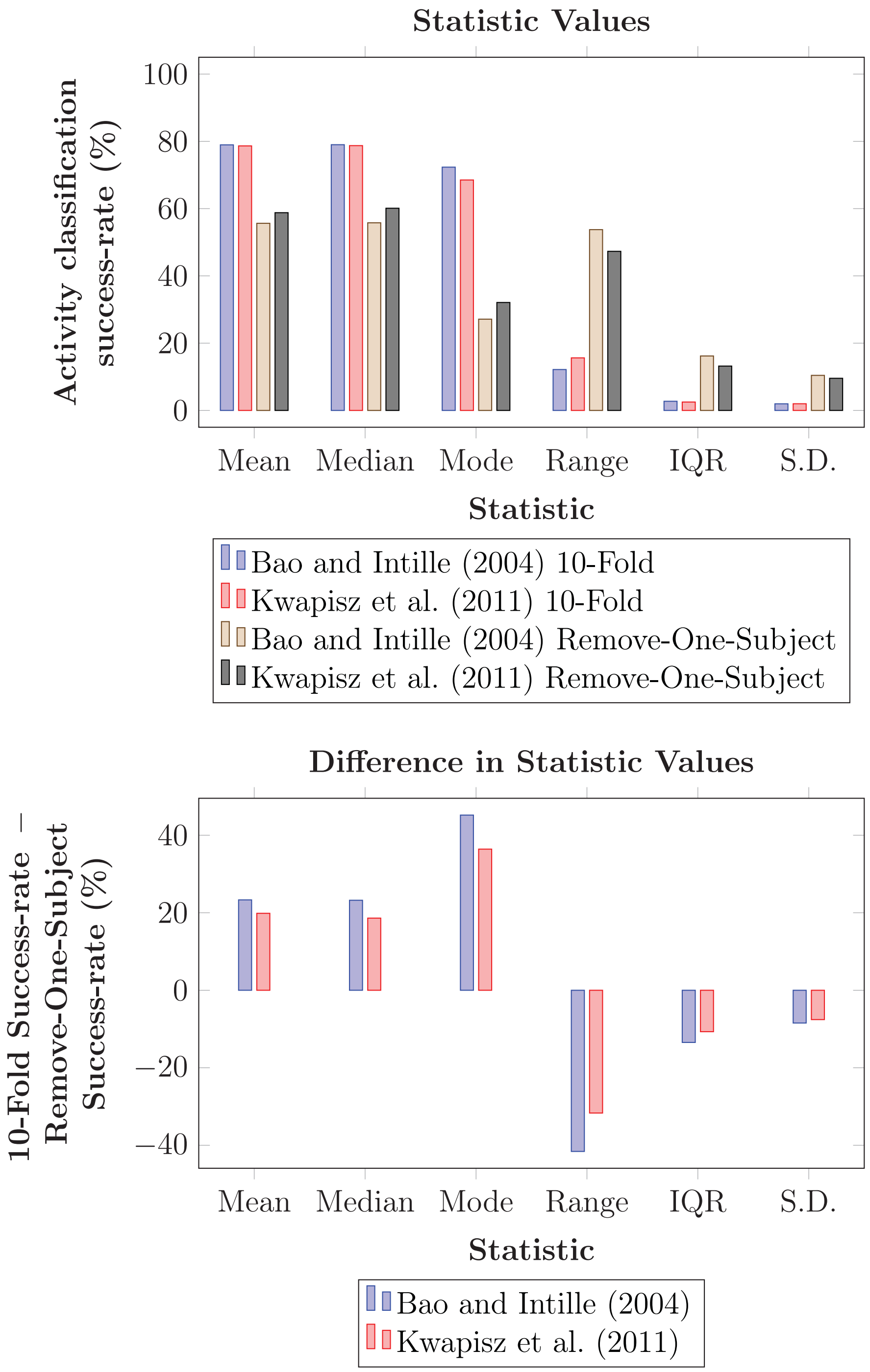


Figure 6.11: Mean, median, mode, range, inter-quartile range (IQR) and standard deviation (S.D.) of overall success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set and using 10-fold cross-validation and remove-one-subject cross-validation (upper). Below is the differences between the statistics obtained from remove-one-subject cross-validation from those obtained from 10-fold cross-validation using Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

Figure 6.12: Boxplots of individual activity success-rates obtained from by using Bao and Intille's feature-set and performing 10-fold cross-validation (left) and remove-one-subject cross-validation (right). The two boxplots share the y-axis on the left.

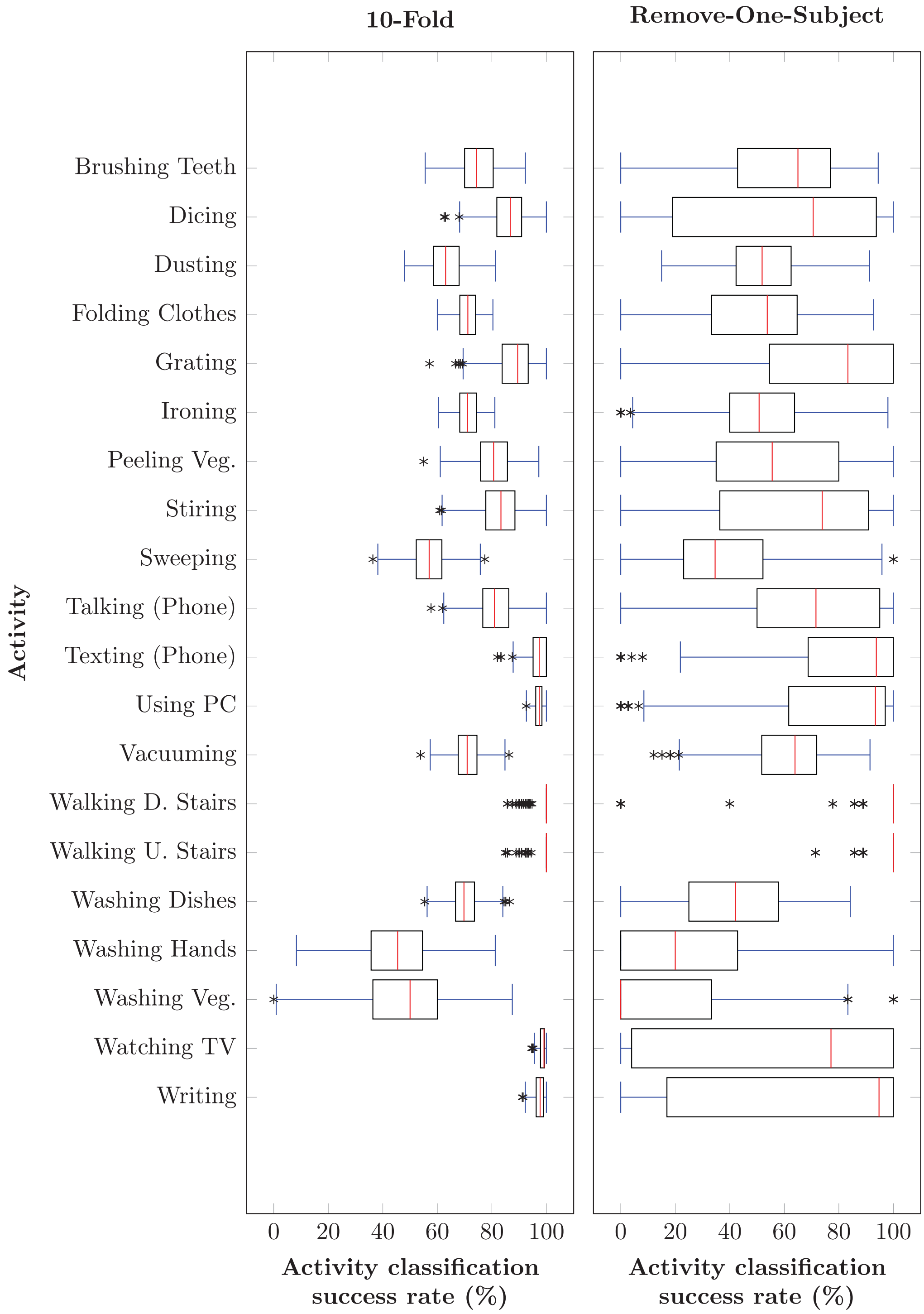


Figure 6.13: Boxplots of individual activity success-rates obtained from by using Kwapisz et al.'s feature-set and performing 10-fold cross-validation (left) and remove-one-subject cross-validation (right). The two boxplots share the y-axis on the left.

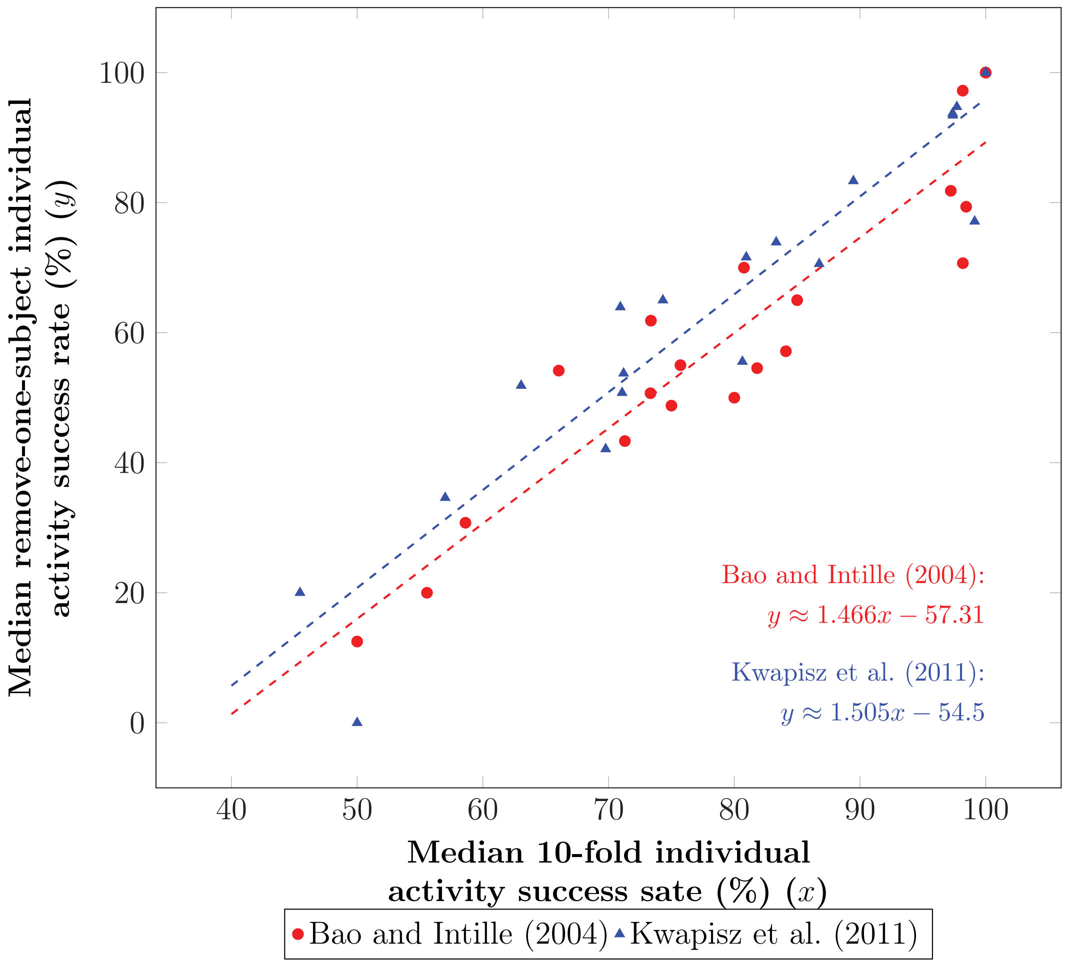


Figure 6.14: The medians of individual activity success-rates obtained using remove-one-subject cross-validation (y) plotted against the medians of individual activity success-rates obtained using 10-fold (x) cross-validation showing strong correlation between the two sets of success-rates. The correlation coefficient between x and y success-rates obtained using Bao and Intille's feature-set is 0.9342, while the correlation coefficient between x and y success-rates obtained using Kwapisz et al.'s feature-set is 0.9473.

tion. The correlation coefficient between the medians of 10-fold cross-validation activity success-rates and the medians of remove-one-subject cross-validation activity success-rates are 0.9342 and 0.9473 for success-rates obtained using Bao and Intille's feature-set and success-rates obtained using Kwapisz et al.'s feature-set respectively. The lines of best fit are observed to be equation 6.2 and equation 6.3, for results from Bao and Intille's feature-set and from Kwapisz et al.'s feature-set respectively.

$$y \approx 1.466x - 57.31 \tag{6.2}$$

$$y \approx 1.505x - 54.5 \tag{6.3}$$

$$\begin{aligned} \text{where, } x &= \text{Median of 10-fold activity success-rates} \\ y &= \text{Median of remove-one-subject activity success-rates} \end{aligned}$$

From equation 6.2 and equation 6.3, we can gather that the minimum activity success-rates obtained using 10-fold cross-validation are higher than those obtained for remove-one-subject cross-validation, however activity success-rates obtained from remove-one-subject cross-validation have a larger spread (about 1.5 times) than those of 10-fold cross-validation.

Discussion

In this section, success-rates obtained using 10-fold cross-validation were compared to those obtained using remove-one-subject cross-validation for both feature-sets studied.

The first research question posed in this section is on how different success-rates obtained from 10-fold cross-validation are from those obtained from remove-one-subject cross-validation. Since it is not possible to directly compare differences in success-rates obtained in each fold, differences in the distributions of the success-rates were used. Three main differences where highlighted: differences in centroids, differences in dispersion, and how well the data fit a normal distribution.

The centroids of success-rates obtained from 10-fold cross-validation were found to be higher than those obtained from remove-one-subject cross-validation. This was observed for both overall success-rates and individual activity success-rates.

The dispersion of success-rates obtained from 10-fold cross-validation were found to be lower than those obtained from remove-one-subject cross-validation. Again, this was observed for both overall success-rates and individual activity success-rates.

When the activity success-rates were plotted against each other, it was observed that median activity success-rates obtained from 10-fold cross-validation strongly correlated with median activity success-rates obtained from remove-one-subject cross-validation.

The results obtained in this section imply that the relative activity classification performance of individual activities using 10-fold cross-validation and remove-one-subject cross-validation are similar. However, the activity classification success-rates obtained from 10-fold cross-validation are higher, less dispersed and require fewer samples to fit a normal distribution.

The drop in success-rates from the success-rates obtained using 10-fold cross-validation to those obtained while using remove-one-subject cross-validation can be explained in terms of inter-subject variability. Differences in motion patterns between subjects result in sensor signal patterns captured from one subject being (to some extent) different from those captured from another subject. Sensor signal patterns of the same subject but different points in time would also exhibit differences, however, the differences are smaller than those between subjects.

When the recorded sensor data is used to train a classifier, the training model matches the testing model more closely data if the data of the same subjects was present in the training set as well as the testing set (as is the case in 10-fold cross-validation). Hence, resulting in better success-rates.

However, the data was used to train the classifier and the data that was used to test the classifier came from different subjects (as is the case in remove-one-subject cross-validation). In this case, the similarity between the training set and the testing set is primarily due to similarities in motions between different subjects (inter-subject similarity).

This has two main implications: more subjects are required to resolve differences between subjects; and the use of the end-user's data results in the activity recognition system having a better activity recognition success-rate.

This analysis used 20 subjects. This is high compared to some of the other studies performed in activity recognition and covered in the activity recognition literature. The difference between success-rates obtained using 10-fold cross-validation and those obtain using remove-one-subject cross-validation suggest that 20 subjects is insufficient to capture the inter-subject variability that exists between subjects performing Activities of Daily Living. Future work should aim to use data from more than 20 subjects.

The results from the analysis also implies that better success-rates can be achieved by activity recognition systems by allowing the activity recognition system the ability to include the data of the end user into it's models. The benefits of this are likely to be higher for systems trained with fewer subjects and lower for systems trained with more subjects.

6.5 Conclusion

In this chapter, the activity success-rates, mutual confusion between activities and success-rates obtained using 10-fold cross-validation and remove-one-subject cross-validation were analysed.

First, the activity success-rates were computed for a number of result sets. The result sets included results from both feature-sets and from both the 3 monitor setup and the 6 monitor setup. The activity success-rates were then compared between each pair of activities within a result set. The comparison was used to create rankings of activities based on the result sets. Finally, similarities in the rankings obtained from all result sets were obtained.

It was observed that using the PC, watching TV, writing and walking(on flat surfaces) are likely to have higher success-rates than any other activity for all result sets analysed. Running was observed to have higher success-rates than these four activities, but this observation was made from the 3 monitor setup result sets only. Walking on a flat surface was observed to have similar success-rates to using the PC, watching TV and writing, but again this observation was made from the 3 monitor setup result sets only. The activities likely to have the next highest success-rates are observed to be texting on the phone, walking down stairs and walking up stairs.

The activities likely to have the lowest success-rates of all activities were observed to be washing vegetables and washing hands. The activities likely to have the next lowest success-rates were observed to be dusting and sweeping, followed by vacuuming, ironing, washing dishes and folding clothes.

A weak correlation was observed between mean success-rates and standard deviation of activities for all result sets. On deeper analysis, it was observed that the weak correlation was a result of there being three distinct groups of activities in all result sets analysed: those with low mean and high standard deviation in success-rates; those with high mean success-rates and low standard deviation in success-rates; and those with a medium mean and standard deviation of success-rates.

The group with the low mean and high standard deviation was observed to consist of washing hands and washing vegetables. The group with high mean success-rates and standard deviations was observed to include walking up stairs, walking down stairs, watching TV, writing, using a PC, texting on the phone and running (using the 3 monitor setup). Other activities fit in the group with medium mean success-rates and medium standard deviation of success-rates.

The results imply that activities like running, using a PC, watching TV, writing and walking on a flat surface are easier to identify and hence activity recognition systems are likely to have higher success-rates while recognising these activities. Activities like washing vegetables and washing hands are more difficult to identify, hence better features are required to more accurately identify these activities.

Next, the mutual confusion between activities was studied so as to identify which activities are highly confusable within the activities and feature-sets studied. It was observed that, for both feature-sets using either the 3 monitor setup or the 6 monitor setup and using any of the *sources* individually or all three combined, the highest mutual confusion errors were between sweeping and vacuuming. The next highest were either between folding clothes and ironing clothes, or washing dishes and washing hands.

In addition, three groups of activities were observed to be highly confusable, for both feature-sets either using 3 monitor setup or 6 monitor setup and using any of the *sources* individually or all three combined: sweeping, vacuuming and dusting; folding clothes and ironing clothes; and washing dishes, washing hands and washing vegetables.

Between result sets computed from individual *sources*, a decreased capability to distinguish between using PC, writing and watching TV is observed in result sets computed from rotational velocities with either feature-sets. It is hypothesised that this is due to these activities being highly stationary and the inability of rotational velocities to be used to distinguish between subject postures.

Finally, success-rates obtained from 10-fold cross-validation were compared to those obtained from remove-one-subject cross-validation. It was observed that distributions of success-rates obtained from 10-fold cross-validation have higher centroids and smaller dispersions than those obtained from remove-one-subject cross-validation. In addition, it was also observed that success-rates obtained from 10-fold cross-validation required fewer samples to fit a normal distribution than success-rates obtained from remove-one-subject cross-validation.

Comparing the medians of activity success-rates obtained from 10-fold cross-validation to medians of activity success-rates obtained from remove-one-subject cross-validation, it was observed that a strong correlation existed. A linear relationship was observed to exist between medians of activity success-rates obtained from 10-fold cross-validation to medians of activity success-rates obtained from remove-one-subject cross-validation, implying that although differences exist in the distribution of overall success-rates between the two cross-validation strategies, the relative performance of individual activities obtained using 10-fold cross-validation was similar to obtained using remove-one-subject cross-validation.

The drop of activity classification success-rates between those obtained from using 10-fold cross-validation to those obtained using remove-one-subject cross-validation is explained to be due to inter-subject variability. While some similarity exists between data captured from one subject to another, a closer similarity exists between data captured from the same subject at two different points in time. Hence, including a subjects data in both the training set and the testing set results in higher similarity between the two sets and hence a higher success-rate. When the training set is created from data captured from different subjects from those used to create the testing set, the success-rates obtained are due to similarities in the motions of different subjects.

This implies that more subjects are needed to catter for these differences due to inter-subject variability. The analysis performed used data captured from 20 subjects. The difference between success-rates obtained using 10-fold cross-validation and those obtain using remove-one-subject cross-validation suggest that 20 subjects is insufficient to capture the inter-subject variability that exists between subjects performing Activities of Daily Living. Hence, future work should aim to use data from more than 20 subjects.

In addition, the results imply that better success-rates can be achieved by activity recognition systems by allowing the activity recognition system the ability to include the data of the end user into it's models. The benefits of this are likely to be higher for systems trained with fewer subjects and lower for systems trained with more subjects.

7 Location and orientation independence in smart-phone-based activity recognition

One of the recent directions of activity recognition research is smart-phone-based activity recognition. The idea is to use sensors available on most current commodity smart phones as sensing nodes in an activity-aware system. The concept clearly has merit since smart phones provide a platform that is:

1. **Familiar to the end-user** since the end-users are already in possession of the device and routinely use the device not only to make calls but also to perform other tasks that have come to form part of smart mobile usage.
2. **Convenient** since the end-users are already in possession of smart phones and make use of them in their daily lives. Hence there is no need for the end-user to purchase, configure and wear extra hardware for the purpose of activity monitoring.
3. **Easily configurable to a sensing node** since smart phones come with a variety of sensors and an internet connection. Data gathered can easily be uploaded to a remote server for further analysis, storage or presentation on a website. In addition, the researchers can work on the algorithms with a user-base and updates and bug fixes can easily be pushed to the end-users.

In terms of activity recognition data-processing, two challenges of smart-phone-based activity recognition set it apart from general on-body activity recognition:

1. **Location independence**: For a smart-phone to be used as a sensing node for activity recognition, the system needs to be able to recognise activities when provided data from any one of multiple body-locations at which phones are carried without requiring the user to carry the phone in a particular body-location or provide input as to which body-location the phone is currently carried in. Since activities have different motions in different locations, this requires that the system be trained using activity data from multiple body-locations.
2. **Orientation independence**: Similarly, the algorithms need to be able to robustly work with data that has been captured in any orientation of the phone. While a wearable sensor can be designed to specificly be worn in a particular orientation relative to a particular body-location, a mobile phone can be carried in any orientation. The orientation of the sensor can impact the data, such that data of the same motion gathered in different orientations can be very different.

An ideal smart-phone-based activity classifier needs to deal with both challenges. But to gain a better understanding, each challenge is studied individually within this chapter. This is achieved by assuming the orientation of the monitor relative to the body-location is fixed while studying location-independence. Similarly, the body-location on which the monitor is mounted is assumed to be known while studying orientation-independence.

In section 7.1, location independence is studied. A fundamental research question is asked: whether the body-location on which a monitor is mounted can be identified without knowing which activity the subject is performing. Next, the success-rates of identifying the body-location are evaluated and compared with reference to the subject's activity. In addition, an evaluation is performed to find out whether the success-rates of identifying the body-location the monitor is mounted on are dependent on the orientation of the monitor.

In section 7.2, orientation independence is studied. The impact of random rotations of the monitor on the activity classification success-rates is studied. In addition, the impact of methods that reorient the data gathered by the monitor to world coordinates on activity classification success-rates is analysed.

The chapter is divided into sections. Each section contains research questions dealing with a specific area of interest. Each section begins with a discussion on the importance of studies on the area of interest; research questions in the area of interest are then posed together with reasons why we wish to attempt to answer these particular questions; a methodology of answering the research questions is given; the results of the analysis are then provided and illustrated; and finally, the conclusions and implications of the result findings are discussed. At the end of the chapter, the analysis, findings and implications of findings of the chapter are summarised.

7.1 Location independence

In general, methods proposed to cater for location-independence, found in the literature, fall into three categories: recognising the location then applying a location specific activity classifier; using one classifier trained with all locations; and using a bag of classifiers, each classifier trained with one location, then using a meta-classifier to pick out the best classification.

In this section, studies in location independence for smart-phone-based activity recognition are performed. In particular, we are interested in answering the following research questions:

1. **Can the body-locations at which a sensor is mounted on a user be identified without knowing the activity that the user is undertaking?**
   It is interesting to discover whether it is possible to identify the body-location on which the monitor was mounted without knowing the activity. By identify-

ing the body-location, a system can apply location-specific models of activities and possibly achieve better activity classification success-rates.

Smart-phone-based activity recognition systems need to cater for various carry locations. Cui at al. (Cui et al., 2007) gives a detailed analysis of where people prefer to carry their phones. The pants pocket (thigh location) was the most preferred by the male survey participants while the hand bag was the most prefered for female participants. The ability to distinguish which body-location the phone is currently carried on can help to improve activity classification not only by allowing for the use of a specific activity model but also allowing the system to determine when the phone is not in one of the trained carry locations.

For a wearable sensor based activity recognition system, the ability to recognise the location of the monitor could help the system identify cases when the monitor has not been worn on the appropriate body-location. In addition, this ability can allow for monitors that can be worn in any one of many body-locations. Hence allowing for a more convenient system where users do not have to worry about where a monitor needs to be worn and possibly lowering the manufacturing cost of the system since all monitors can be identical to each other.

A previous method proposed by Kunze et al. ((Kunze et al., 2005)) relied on first identifying when a specific activity A is occurring (the activity was walking for Kunze et al.); then identifying the carry location based on location-specific models of the identified activity; then using activity models of the identified carry location to identify all activities till the next identified occurence of activity A. While Kunze et al. have shown that this method works well, it is a complex approach and errors in identifying the carry location persist until the next identification of the carry location.

There has, to date, been no communication in the research literature, concerning the identification of the body-location on which the monitor is mounted on, that is independent of the subject's activity.

2. **Is there an interplay between user activity and body-location that a monitor is mounted at, such that body-locations identified during some activities are more accurate than other activities?**

   Kunze et al.'s algorithm (explained in the previous research question), specificly targets the occurence of walking. One reason given by Kunze et al. is that this is because walking is a common activity in the lives of end users. The other reason is that the motion signature of walking is distinct enough that it can be recognised without any assumption about the location of the monitor location.

   It is likely that the difference of motions between body-locations for some activities is higher than in other activities. The information of which activities result in lower success-rates in identifying the body-location on which the monitor is mounted at could be useful in assigning a likelihood of the body-location classification being true based on the activity the subject is performing.

3. **Are the success-rates of identifying the body-location (at which the monitor is mounted on) dependent on the orientation of the monitor relative to the body-location being fixed?**

   The dependency of identifying the body-location on which the monitor is mounted on, on the fixed orientation of the monitor relative to the body-location, is unclear.

   In order to independently study location independence from orientation independence, an assumption was made that the orientation of the monitor would be fixed relative to the body-location the monitor is mounted on. However, the orientation of a smart-phone may change from time to time. Hence, it is necessary to analyse what impact of this assumption has on the results obtained.

Methodology

In order to analyse how well the body-location on which a monitor was mounted can be identified, a classifier is trained with feature-vectors extracted from all monitors then tested to determine whether the source body-locations of feature-vectors in the testing fold can be identified. Specifically, to answer the third research question in this section, a random rotation was applied within each window so as to eliminate any impact of the orientation of the monitor being fixed. The random rotations were not applied for result sets obtained to answer the other research questions. The process of computing result sets to answer the research questions is hence as follows:

For each monitor, all three *sources* were used. A sliding window of 10 seconds with 50% overlap was used.

Window lengths of 10 seconds were used because 10 seconds was observed to be the maximum window length used in activity recognition literature, having only been used by Kwapisz et al. (2011) and Patel et al. (2009). In the analysis of activity recognition literature performed by Lockhart and Weiss (Lockhart & Weiss, 2014), window lengths reported to have been used in activity recognition literature were observed to have a median of 3 seconds and the maximum window length they observed was 10 seconds.

A 50% window overlap was used because it was observed that 50% window overlaps are common within the literature review having been used by Bao and Intille (2004); Figo et al. (2010); He et al. (2008); Krishnan and Panchanathan (2008); Kunze et al. (2005); Preece, Goulermas, Kenney, and Howard (2009); Ravi et al. (2005); Shoaib et al. (2014) and Sun et al. (2010). However, other window overlaps also exist in the literature review including: no overlap ((Kwapisz et al., 2011)), 20% ((Reiss, 2014)), 25% overlap ((Henpraserttae et al., 2011)), 33% overlap ((Lester et al., 2005)).

Optionally, random rotations were introduced into each window to eliminate any impact of the monitor orientation being fixed. The random rotations are uniformly distributed over all possible rotation axes with rotation magnitudes in the range $[0°, 360°)$. This represents a more extreme situation than could be expected with a loose and floating sensor

The data was downsampled to frequencies ranging from 112Hz to 128Hz in intervals of 1Hz. This is because, as observed in section 4.2, changes in downsampling frequency impact the success-rate. Low pass filtering was performed prior to downsampling.

Features were then extracted as explained in the respective feature-set's paper (Bao and Intille (2004) and Kwapisz et al. (2011), refer to section 3.5 for more details). In addition, a Hamming window was applied to each window before extracting the frequency-domain features in Bao and Intille's feature set.

Finally, classification was performed using the J48 decision tree from the WEKA toolkit. The J48 decision tree is an implementation of the C4.5 algorithm (Hall et al., 2009). The C4.5 decision tree was found to perform best by Bao and Intille (Bao & Intille, 2004), and second best by Kwapisz et al. (Kwapisz et al., 2011).

Experiments were performed using 10-fold cross-validation. The classifier was trained with data from all monitors, then tested to find whether it could identify the body-locations of the feature-vectors in the testing fold.

Figure 7.1 shows an overview of this process.

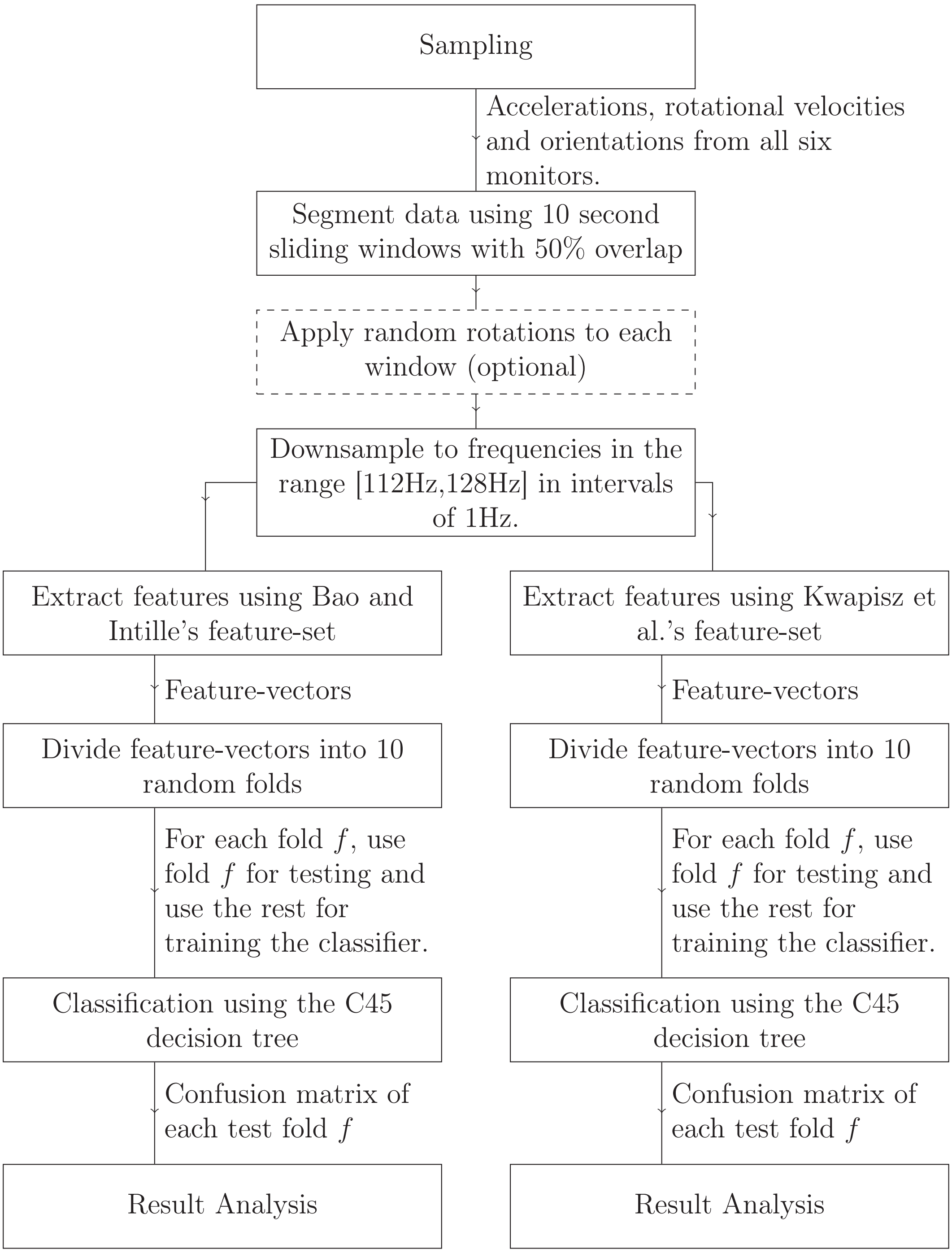
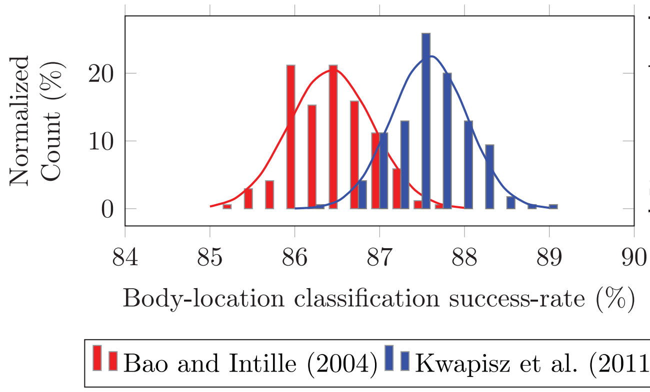


Figure 7.1: Overview of the methodology used to compute result sets used to answer the research question in this section.

Results

Can body-locations be identified without knowing the activity?



	Bao	***Kwapisz***
$\bar{x}$	86.42	87.58
σ	0.8361	0.7906
R^2	0.8726	0.9419
SSE	84.18	42.38

Figure 7.2: Normalised histograms of body-location classification success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set. Classification was performed using all monitors but excluded walking and running activities.

Figure 7.2 shows normalised histograms of body-location classification success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set. From the figure, we can observe that high success-rates are obtained using both feature-sets. The mean body-location classification success-rate is observed to be 86.42% for Bao and Intille's feature-set and 87.58% for Kwapisz et al.'s feature-set. Moreover, the standard deviation of the body-location classification success-rates are observed to be low which implies that the high success-rates are obtained consistently from all testing folds.

One-tailed one sample t-tests with $\alpha = 0.05$ were run on the result sets shown in figure 7.2 to test the null hypothesis that the result sets came from normal distributions with mean of chance (i.e. 16.67%) and unknown variance. Tests rejected the null hypothesis for each of the result sets implying that the results obtained are statistically significantly above chance.

Figure 7.3 shows a box-plot of the body-location classification success-rates obtained in identifying each body-location using each feature-set.

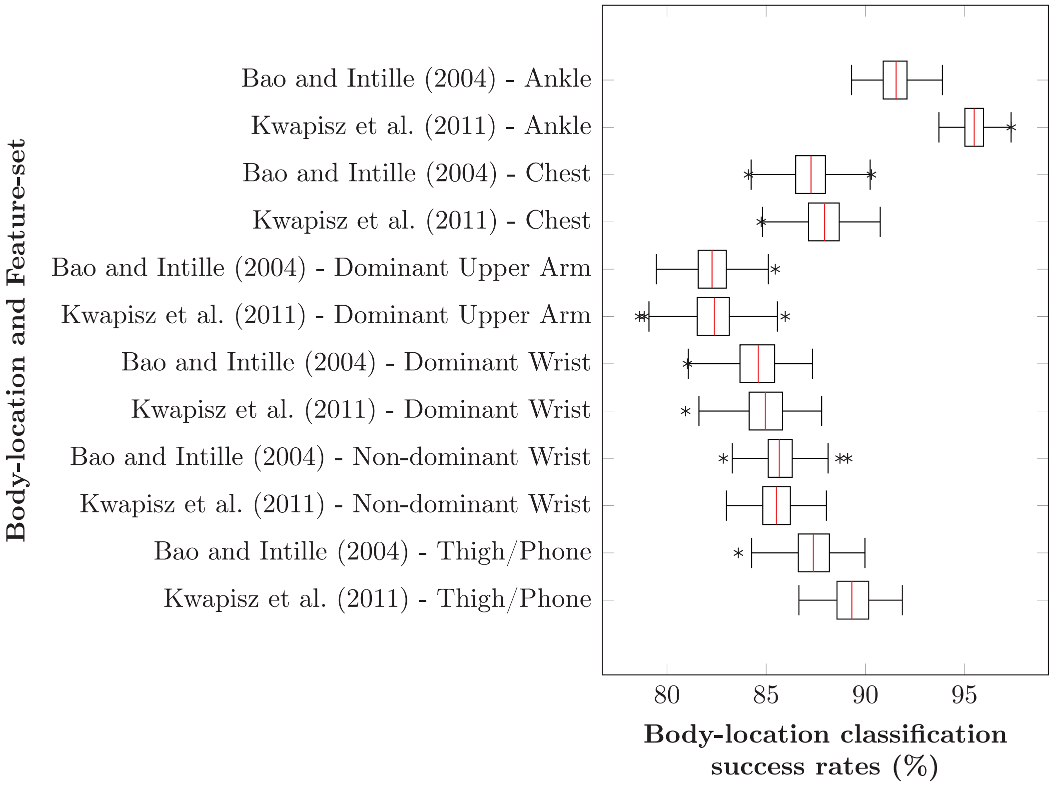


Figure 7.3: Boxplot of the body-location classification success-rates obtained in identifying each body-location using each feature-set.

Two-tailed two sample t-tests with $\alpha = 0.05$ were run on the success-rates of identifying each pair of body-locations using the same feature-set, to test the null hypothesis that the success-rates of identifying the pair of body-locations came from independent random samples of normal distributions with equal means and unknown variance. The tests rejected the null hypothesis for each of the pairs of body-locations and feature-sets except between thigh/phone and chest body-locations using Bao and Intille's feature-set. This implies that there is insufficient evidence to support the hypothesis that the success-rates of identifying the thigh/phone and those of identifying the chest while using Bao and Intille's feature-set are different. However, there is sufficient evidence to support the hypothesis that the success-rates of identifying other pairs of body-locations using Bao and Intille's feature-set, and all body-locations using Kwapisz et al.'s feature-set, are different.

From the box-plot, it can be observed that the body-location identified with the highest success-rate is the ankle for both feature-sets. The next highest success-rates are observed to be the thigh/phone while using Kwapisz et al.'s feature-set; or either thigh/phone or chest while using Bao and Intille's feature-set. These are followed by the non-dominant wrist, dominant wrist and finally dominant upper arm. The medians of all the result sets are observed to be above 80%.

Is there an interplay between user activity and body-location that a monitor is mounted at, such that body-locations identified during some activities are more accurate than other activities?

Figure 7.4 shows a box-plot of the body-location classification success-rates obtained during each activity using Bao and Intille's feature-set and Kwapisz et al.'s feature-set. From the boxplot, it can be observed that the success-rates are generally high for all activities. The median success-rates range from 97% for writing using Kwapisz et al.'s feature-set to 69% for talking on the phone using Bao and Intille's feature-set. The medians of success-rates between the two feature-sets are noted to be highly similar, having a correlation coefficient of 0.9255.

One-tailed one sample t-tests with $\alpha = 0.05$ were run on each of the individual activity's success-rates obtained from each of the result sets to test the null hypothesis that the result sets came from normal distributions with mean 16.67% (i.e. chance) and unknown variance. Tests rejected the null hypothesis for each individual activity's success-rates obtained from each of the result sets implying that the success-rates of identifying locations obtained for each individual activity from each result set are significantly above chance.

Figure 7.5 shows a bar graph of the mean success-rates (and standard deviations shown using error bars) of the body-location classification success-rates obtained for each activity using Bao and Intille's feature-set and Kwapisz et al.'s feature-set. From the bar graph, it can be observed that the highest success-rates in identifying body-locations are obtained during the activity writing. Similar high success-rates

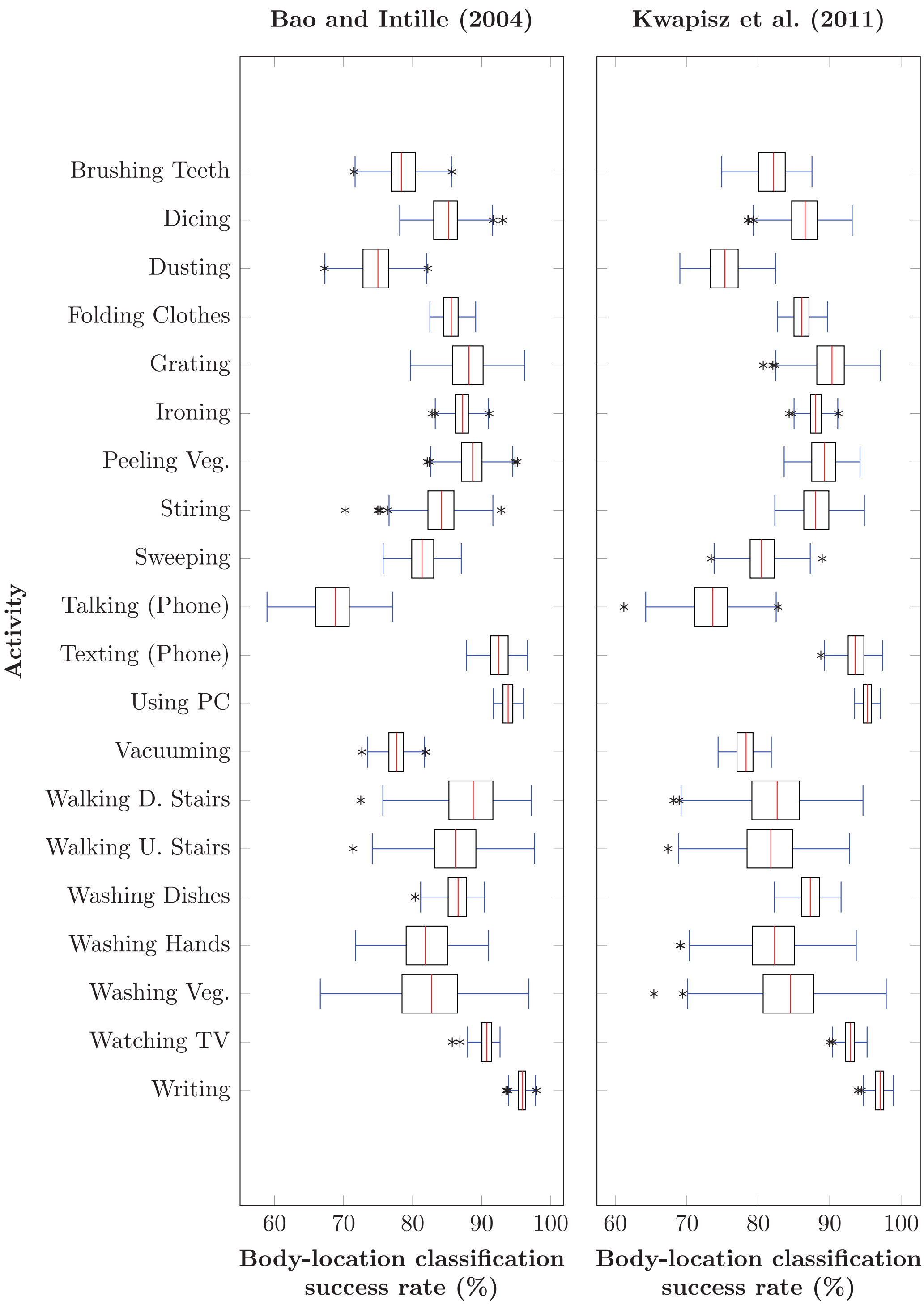


Figure 7.4: Boxplots of the body-location classification success-rates obtained for each activity using Bao and Intille's feature-set (left) and Kwapisz et al.'s feature-set (right). The two boxplots share the y-axis on the left.

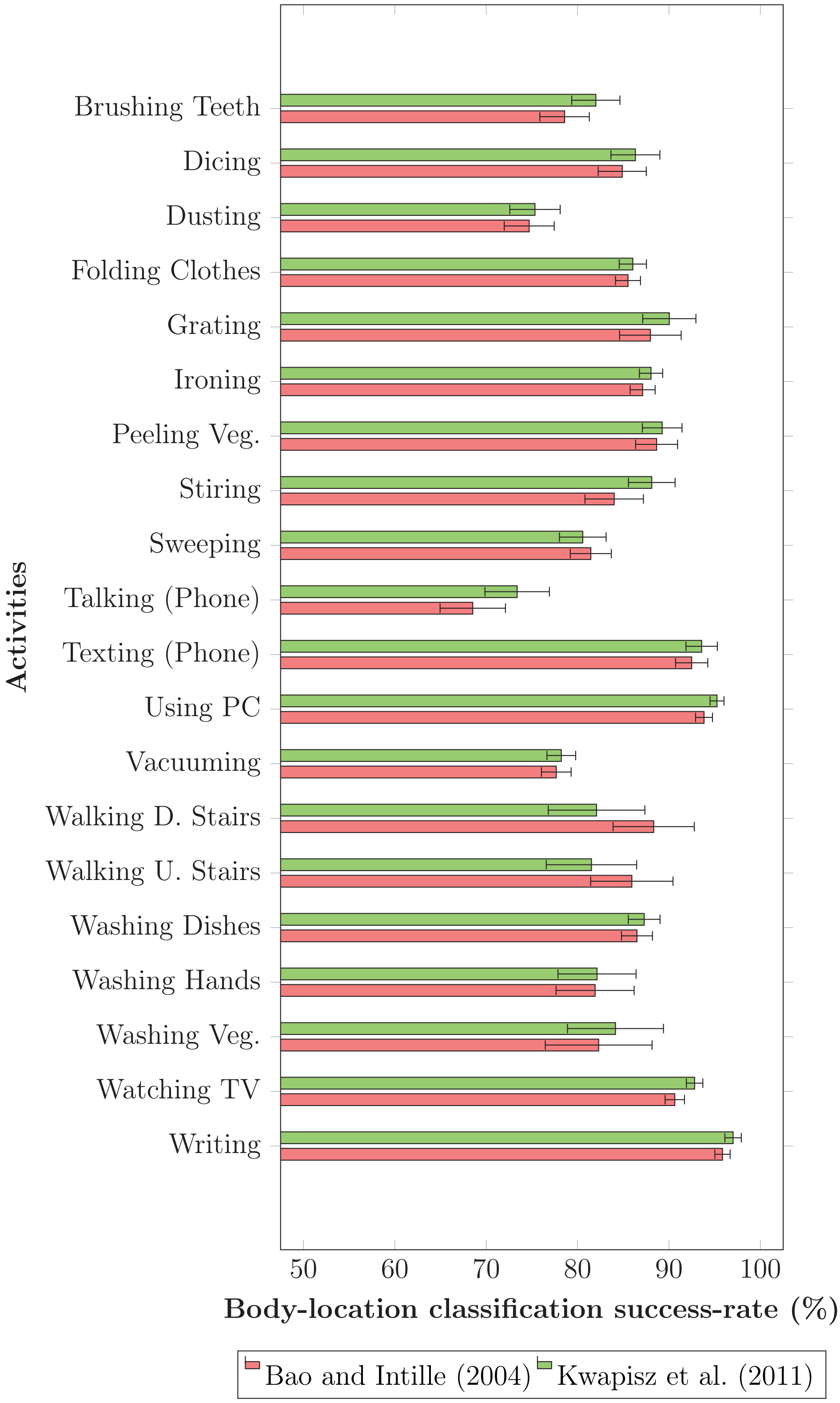


Figure 7.5: Mean success-rates (and standard deviations shown using error bars) of the body-location classification success-rates obtained for each activity using Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

can be obtained while using the PC, texting on the phone and watching TV. The lowest success-rates are obtained while talking on the phone.

The ranking of activities based on the likelihood of identifying body-locations at a higher success-rate during one activity than during other activities is presented in figure 7.6 for Bao and Intille's feature-set and figure 7.7 for Kwapisz et al.'s feature-set. From the two figures, it can be observed that the highest success-rates of identifying body-locations are likely to be obtained while (in decreasing order) writing, using a PC, texting on the phone and watching TV, for both feature-sets. Alternatively, the lowest success-rates of identifying body-locations are likely to be obtained while (in increasing order) talking on the phone, dusting and vacuuming, for both feature-sets. The rankings of other activities differ depending on the feature-set used.

Are the success-rates of identifying the body-location (on which the monitor is mounted on) dependent on the orientation of the monitor relative to the body-location being fixed?

Figure 7.8 shows normalised histograms of body-location classification success-rates obtained from using Bao and Intille's feature-set and Kwapisz et al.'s feature-set that had random rotations added to each window of data. From the figure, we can observe that the success-rates obtained by both feature-sets are significantly lower than those obtained without random rotations (shown in figure 7.2). The mean body-location classification success-rate is observed to be 45.66% for Bao and Intille's feature-set and 54.85% for Kwapisz et al.'s feature-set as opposed to 86.42% and 87.58% without rotations.

One-tailed one sample t-tests with $\alpha = 0.05$ were run on each of the result sets that were computed with randomised rotations to test the null hypothesis that the result sets came from normal distributions with a mean of chance (i.e. 16.67%) and unknown variance. Tests rejected the null hypothesis for each of the result sets

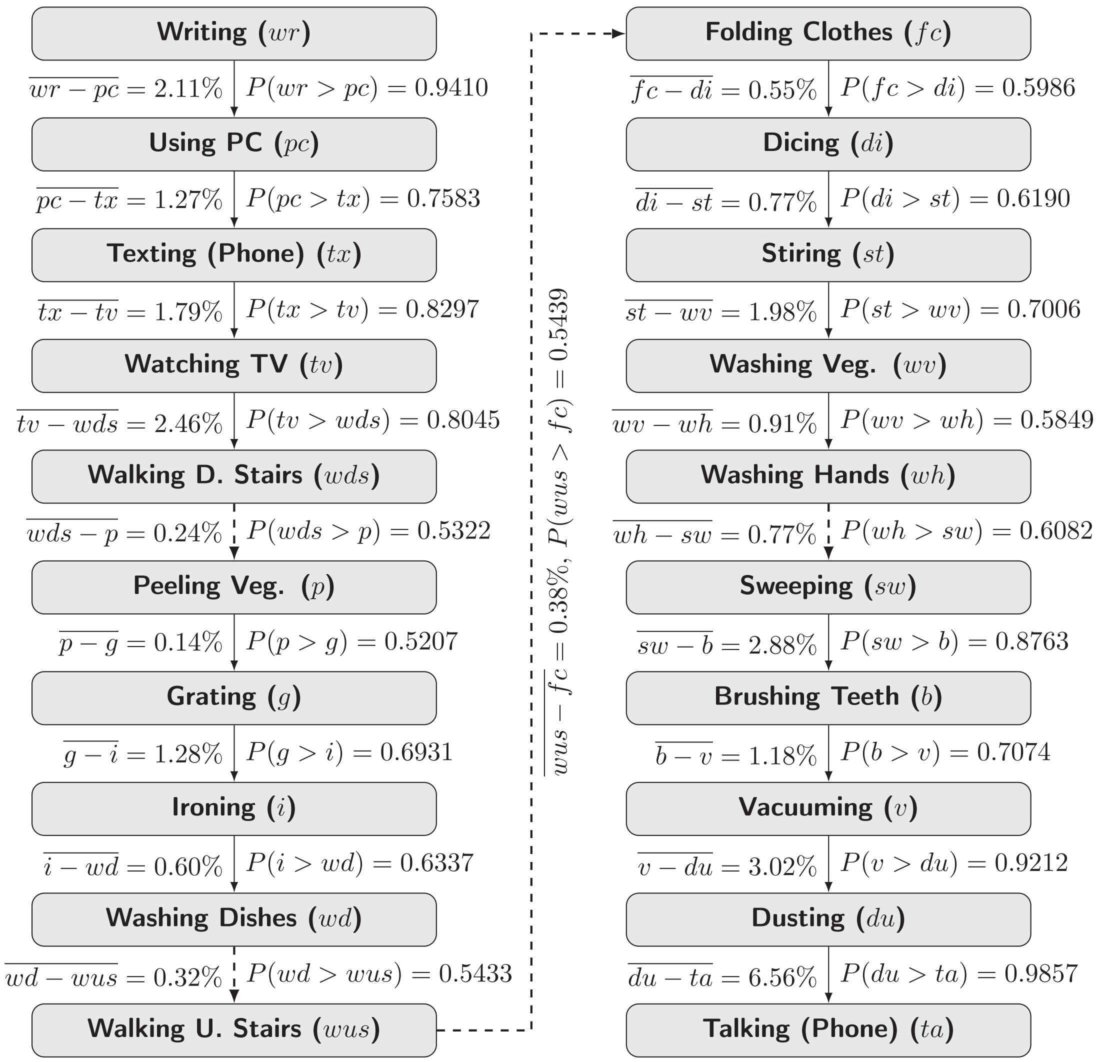


Figure 7.6: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates at identifying body-locations while performing one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. Activities without a statistically significant difference in success-rates are separated by dashed lines. The results are generated from data extracted using Bao and Intille's feature-set.

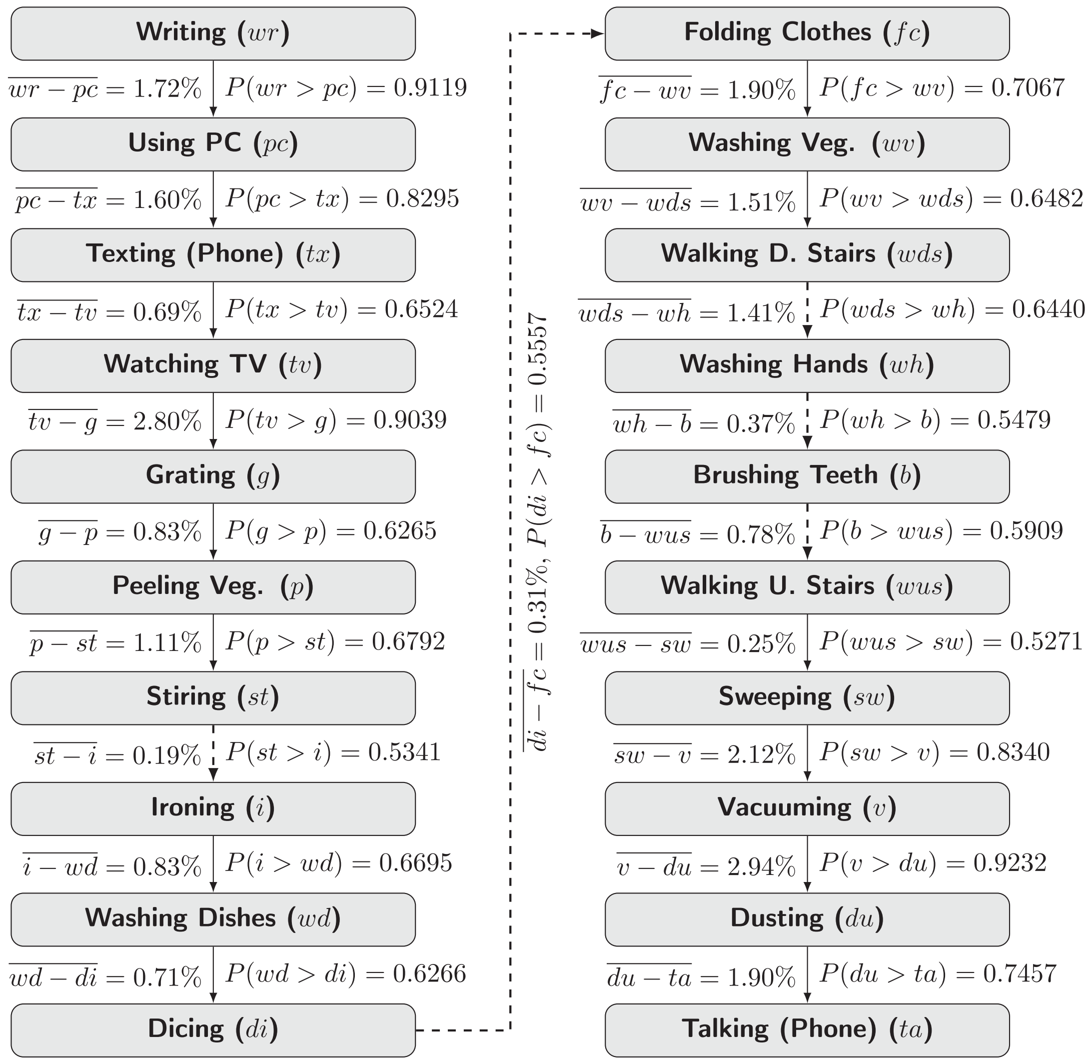
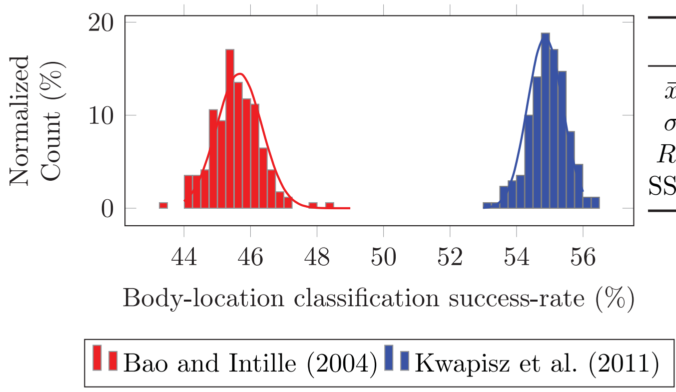


Figure 7.7: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates at identifying body-locations while performing one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. Activities without a statistically significant difference in success-rates are separated by dashed lines. The results are generated from data extracted using Kwapisz et al.'s feature-set.



	Bao	***Kwapisz***
$\bar{x}$	45.66	54.85
σ	0.9826	0.8641
R^2	0.9301	0.9835
SSE	29.31	9.32

Figure 7.8: Normalised histograms of body-location classification success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set with random rotations introduced into each window so as to remove the impact of monitor orientation. Classification was performed using all monitors but excluded walking and running activities.

implying that the results obtained are significantly above chance.

A two-tailed two sample t-tests with $\alpha = 0.05$ between the result sets obtained from data with random rotations and result sets obtained from data without random rotations of each feature-set rejected the null hypothesis that the result sets came from independent random samples from normal distributions with equal means. Hence, the results set obtained from data with random rotations were found to be statistically different from those obtained from data without random rotations.

Figure 7.9 shows a box-plot of the relative decrease in body-location classification success-rates due to the introduction of random rotations into windows. The relative decrease was computed using equation 7.1. From figure 7.9, it can be observed that the average relative decrease is between 40% and 50%.

$$\text{Relative decrease} = \frac{S - S_{wRot}}{S}\,100 \tag{7.1}$$

where, S = Test-fold success-rate without random rotations ,

S_{wRot} = Test-fold success-rate with random rotations

In fact, the mean relative decreases are 47.32% and 37.85% and the standard

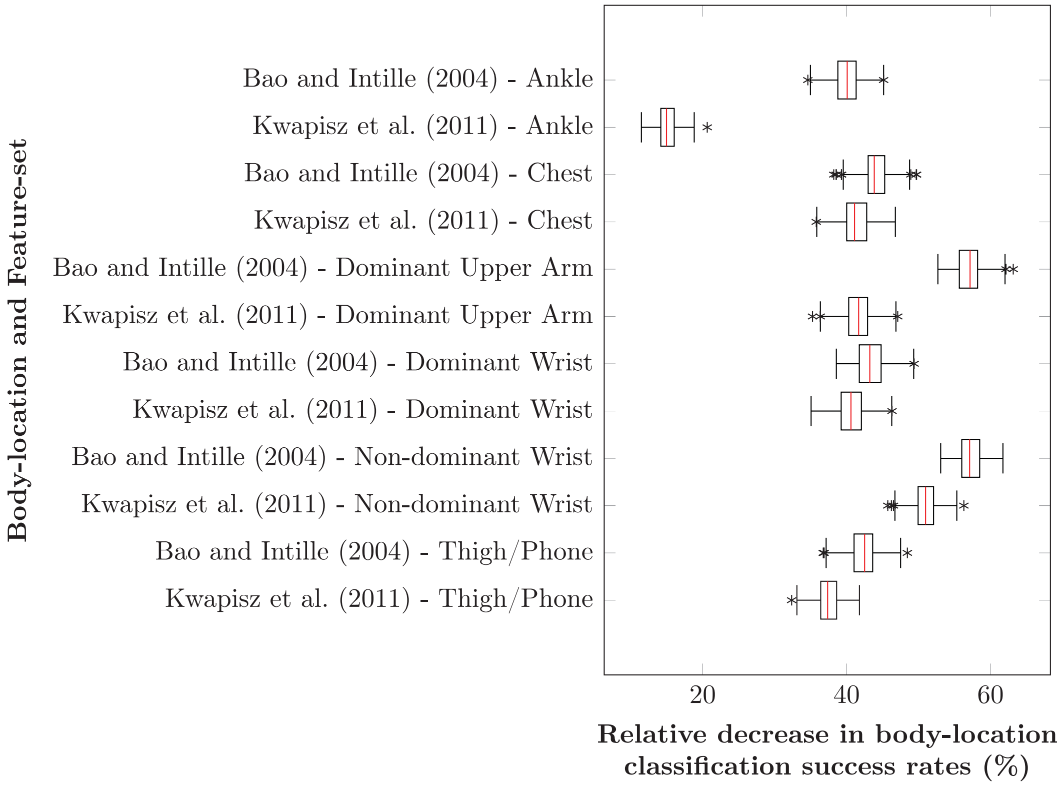


Figure 7.9: Boxplot of the relative decrease in body-location classification success-rates due to the introduction of random rotations into windows. Decrease in body-location classification success-rates observed for both feature-sets and for all six body-locations studied are displayed.

deviations are 7.28% and 11.13%, for the two feature-sets respectively. Hence, it can be said that the addition of random rotations to each window resulted in a significant drop in success-rates.

The result set that was impacted the least is observed to be that of identifying the ankle body-location using Kwapisz et al.'s feature-set. This result set is also the result set that had the highest success-rate in identifying body-locations (see figure 7.3). The result set that was impacted the most is observed to be that of identifying the non-dominant wrist using Kwapisz et al.'s feature-set and either the dominant upper arm or non-dominant wrist using Bao and Intille's feature-set.

Figure 7.11 shows the mean and standard deviations of body-location classification success-rates obtained for each activity using Bao and Intille's feature-set and

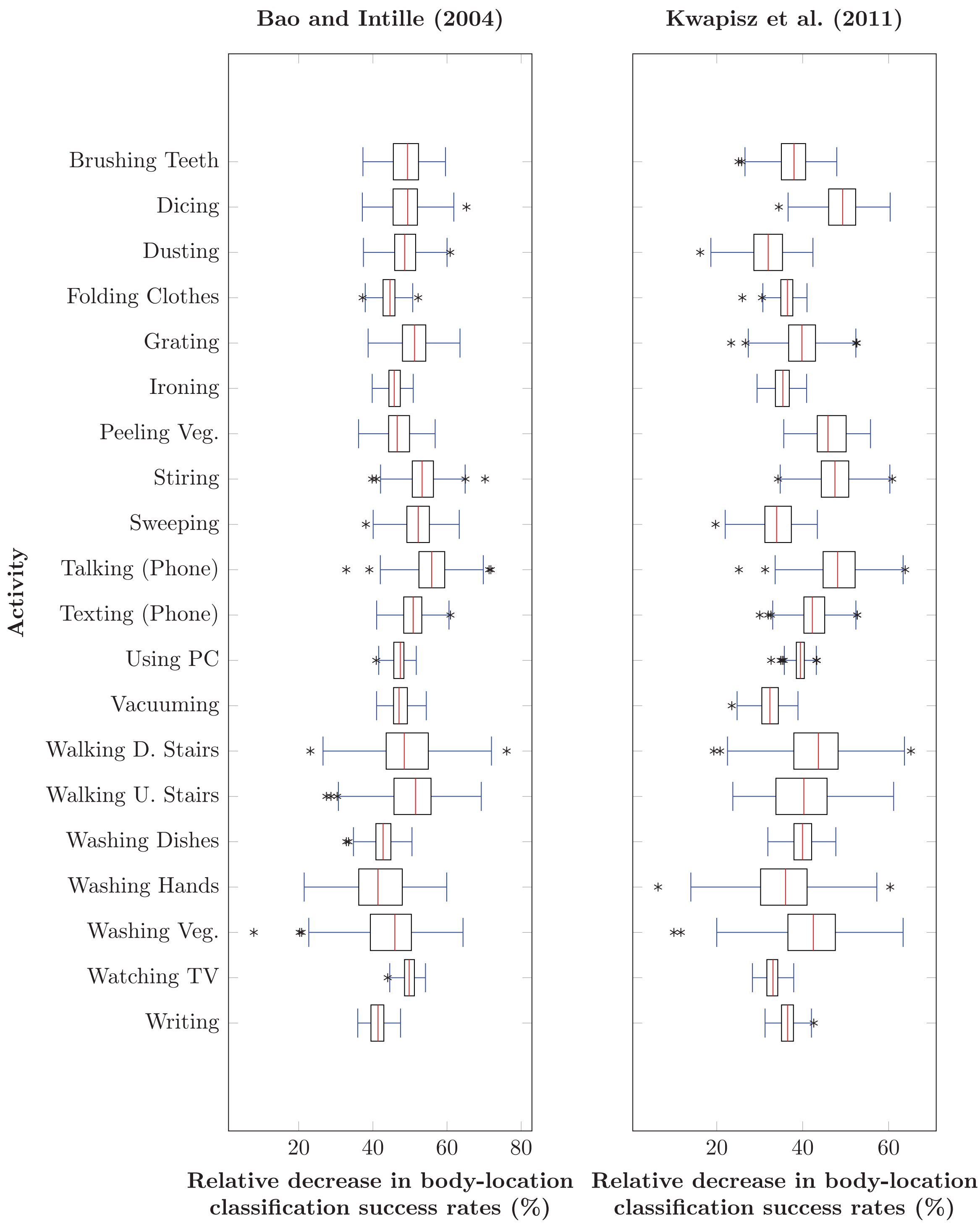


Figure 7.10: Boxplots of the relative decrease in body-location classification success-rates due to the introduction of random rotations into windows. Relative decreases in body-location classification success-rates obtained by using both Bao and Intille.'s feature-set (left) and Kwapisz et al.'s feature-set (right) are shown. The two boxplots share the y-axis on the left.

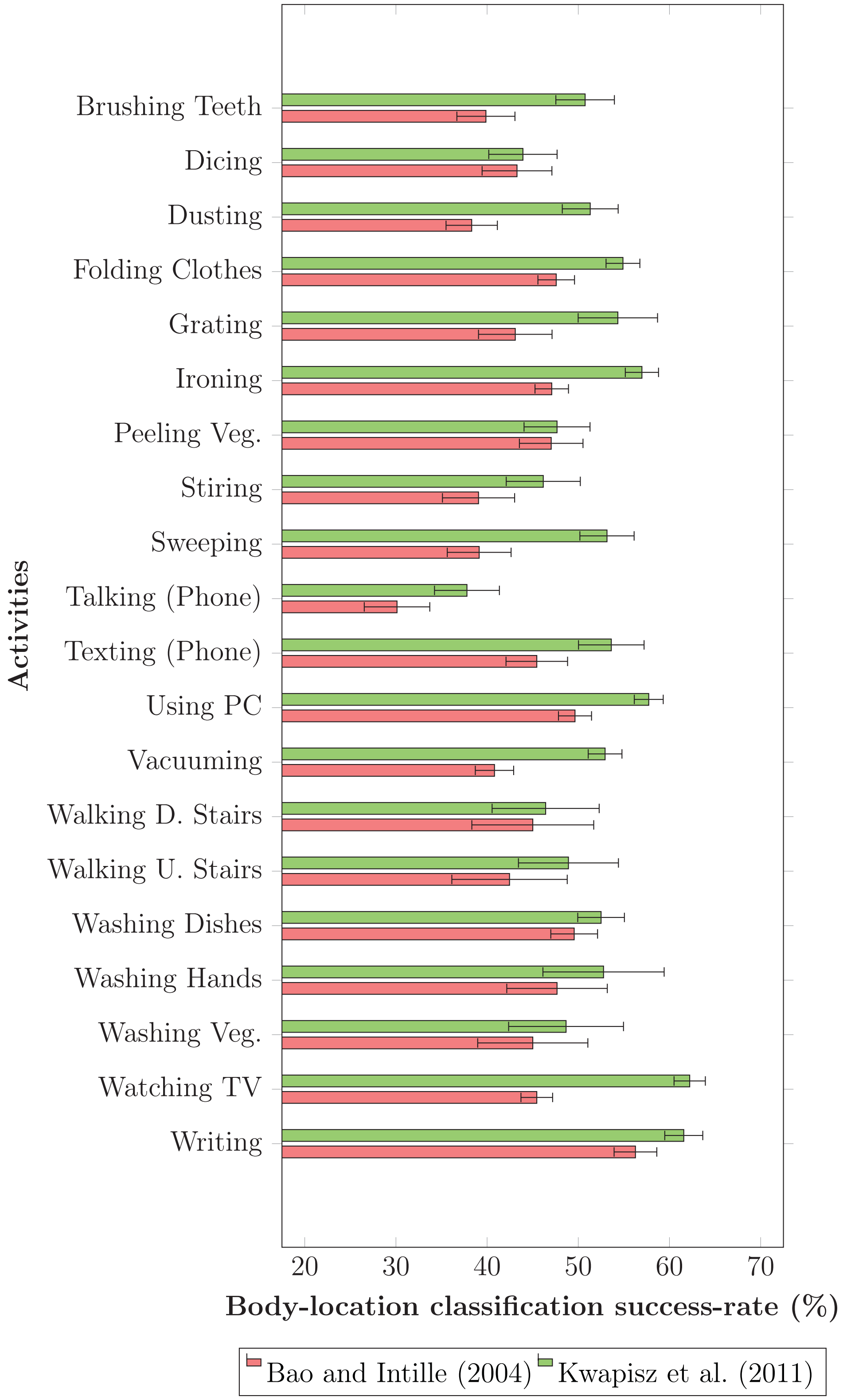


Figure 7.11: Mean success-rates (and standard deviations shown using error bars) of body-location classification success-rates obtained for each activity using Bao and Intille's feature-set and Kwapisz et al.'s feature-set when applied to data with random rotations.

Kwapisz et al.'s feature-set when applied to windows that have random rotations. From the figure, it can be observed that Bao and Intille's feature-set results in lower success-rates of identifying body-locations for all activities. Compared to the performances of activities without random rotations (see figure 7.5), activities like talking on the phone are still observed to perform poorly compared to other activities, while activities like writing and watching TV are still observed to perform well.

The ranking of activities based on the likelihood of identifying body-locations at a higher success-rate during one activity than during other activities and using data that has random rotations applied to it, is presented in figure 7.12 for Bao and Intille's feature-set and figure 7.13 for Kwapisz et al.'s feature-set. From the two figures, it can be observed that, for both feature-sets, the lowest success-rates of identifying body-locations are likely to be obtained while talking on the phone if the data has random rotations applied. The ranking of other activities varies depending on the feature-set used.

Discussion

First, feature-vectors were extracted using Bao and Intille's feature-set and using Kwapisz et al.'s feature-set from data captured from each body-location. The feature-vectors were then classified to as to identify the body-location on which the monitor was mounted on.

The results obtained showed high success-rates in identifying the body-locations for both feature-sets. One-tailed one sample t-tests with $\alpha = 0.05$ confirmed that the success-rates were significantly above chance.

A deeper analysis of the obtained results showed that the most easily identified body-location using either feature-set was the ankle; followed by either the chest or thigh/phone depending on the feature-set; the non-dominant wrist; dominant wrist; and finally the dominant upper arm.

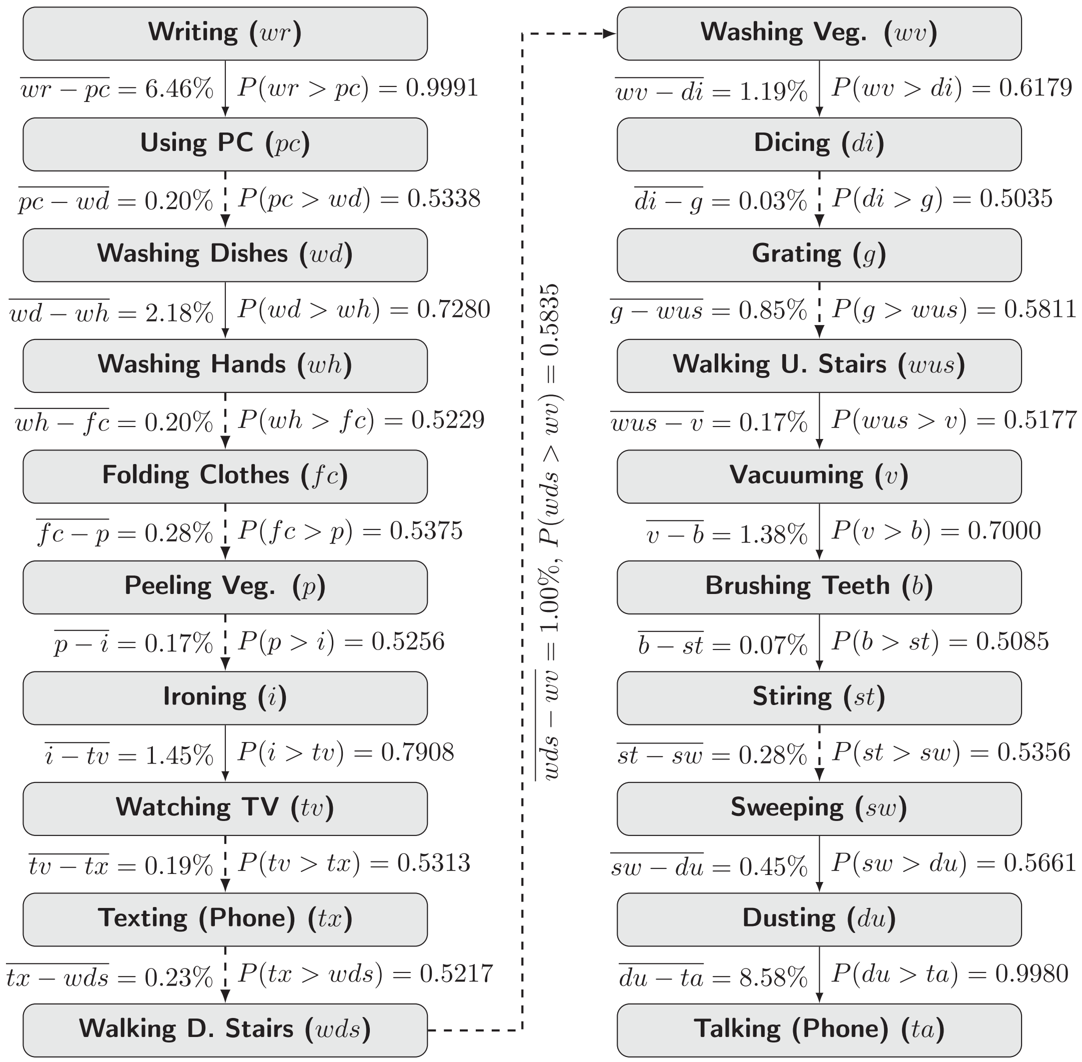


Figure 7.12: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates at identifying body-locations while performing one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. Activities without a statistically significant difference in success-rates are separated by dashed lines. The results are generated from data with random rotations extracted using Bao and Intille's feature-set.

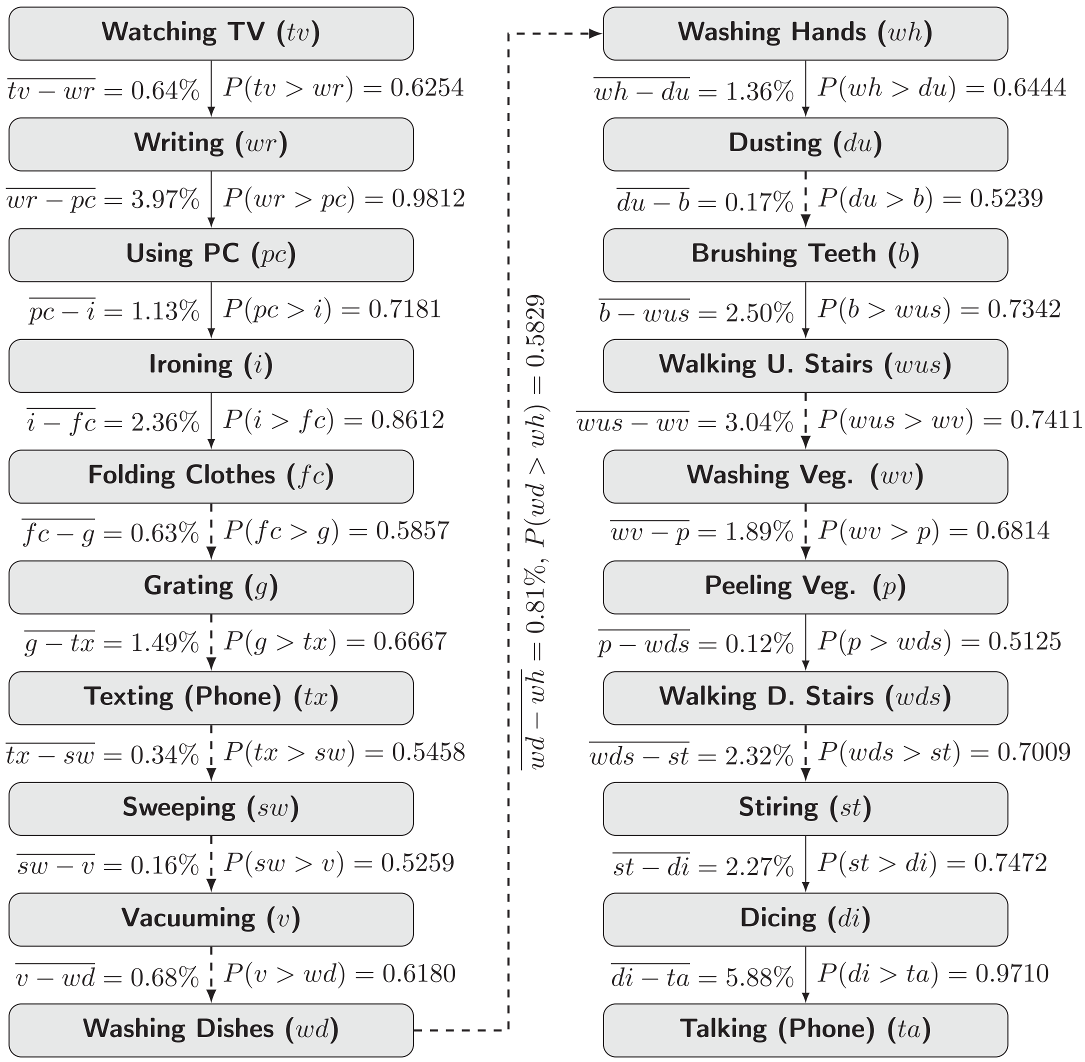


Figure 7.13: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates at identifying body-locations while performing one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. Activities without a statistically significant difference in success-rates are separated by dashed lines. The results are generated from data with random rotations extracted using Kwapisz et al.'s feature-set.

Next, the success-rate of identifying each body-location as a function of the subject's activity was studied. It was observed that while success-rates of identifying body-locations are generally significantly higher than chance for all activities, the success-rates obtained for some activities were less than those obtained by other activities. It was observed that among the activities that result in higher success-rates in identifying the body-location on which the monitor was mounted for both feature-sets are: writing, using a PC, texting on the phone and watching TV. Among the activities that result in the lowest success-rates in identifying the body-location on which the monitor was mounted for both feature-sets are: talking on the phone, dusting and vacuuming.

The activities that result in higher success-rates are highly stationary. It is possible that the high success-rates obtained are partially due to the highly constant orientation of the monitors. A seemingly counterexample to this preposition is that the activity with the worst success-rate, talking on the phone, is also a highly stationary activity. However, since the subjects were asked to write down a number spoken to them over the phone while recording the talking on the phone conversation, it is possible that the shifting of the phone from the dominant wrist to the non-dominant wrist so that the dominant wrist can be used to write the number down adds confusion to the body-location identification during talking on the phone.

This was tested in the next research question, where random rotations were introduced into each window before extracting feature-vectors using both feature-sets. This eliminates any impact of the monitor orientation being fixed.

The results showed a significant reduction in success-rates in all body-locations and for all activities. However, the success-rates were still observed to be significantly higher than chance. This implies that the success-rates observed in identifying the body-locations the monitors were mounted on, are only partially due to the fixed orientation of the monitors relative to the body-location mounted on.

Hence, the results obtained in this section show that it is possible to identify the body-location a monitor is mounted on without knowing the subject's activity. The success-rates obtained vary depending on the subject's activity, but are higher than chance for all activities. In addition, a significant difference is observed between result sets where the orientation of the monitor relative to the body-location is fixed and result sets where the orientation of the monitor relative to the body-location is transient. However, even when the orientation of the monitor relative to the body-location is transient, the success-rates obtained are still above chance.

This implies that both wearable activity recognition systems and smart-phone-based activity recognition systems can identify the body-location the monitor is mounted on, or the carry location of the phone. This would allow activity recognition systems to apply body-location specific models of activities so as to achieve higher success-rates. In addition, wearable activity recognition systems could be made more convenient for the user by recognising which body-location the monitor is mounted on and hence users do not have to worry about where a monitor needs to be worn and possibly lowering the manufacturing cost of the system since all monitors can be identical to each other.

It is important to note that the results obtained reflect success-rates achieved when only one window is classified. The success-rates are likely to improve significantly by using a model of the system that takes into consideration the impact of time on the state of the system. This can be performed by using Hidden Markov Models or Markov Chains, but would require that the likelihood of the monitor being in different body-locations within the system usage and the likelihood of changes in body-locations on which the monitor is mounted on, be modelled.

7.2 Orientation independence

Methods of dealing with sensor orientation changes, found in the literature, fall into three categories (Henpraserttae et al., 2011): using features that are orientation-independent (e.g. vector magnitudes); applying various data preprocessing methods to reorient the data to world coordinates before extracting features; and training the classifier(s) with data of as many orientations as possible.

Of the three methods of dealing with changes in sensor orientation, it has been shown (Yang, 2009) that reorienting the data to world coordinates results in higher success-rates than using orientation-independent features. The explanation proposed in the literature is that using orientation-independent features (like vector magnitudes instead of the vector itself) discard information (like the vector direction) which could be useful for differentiating between some activities.

In addition, it has also been shown (Henpraserttae et al., 2011) that reorienting the data to world coordinates results in higher success-rates than training the classifier(s) with data of multiple orientations. Training the classifier with different models, each of which includes the same motions but with a different monitor orientation requires more effort in data collection.

Two methods of reorienting data to world coordinates are possible:

Accelerometer-based method : This is the method proposed by Henpraserttae et al. and explained in detail in their paper Henpraserttae et al. (2011). It is based on the method proposed by Yang (2009) which in turn is based on Mizell (2003).

It involves computing the gravity vector as the mean of the accelerations gathered in the window. This gravity vector provides the vertical (up-down) axis. Next, the anteroposterior (forward-backward) axis is obtained as the first principal component of the accelerations projected onto the horizontal plane. The mediolateral (left-right) axis is obtained as the cross-product of the anteroposterior axis and the vertical axis.

Since the gravity vector is computed from the average of the accelerations of the window, the orientation obtained for the window is computed as the mean orientation of the different orientations the monitor was in during the window.

Orientation-based method : This method makes use of the orientation of the monitor provided by the IMU. The IMU computes the orientation by integrating the rotational velocities obtained from the gyroscope then including the gravity vector computed from the accelerations to correct the eventual drift. Hence, the orientation of the monitor relative to world coordinates at each sample is known.

To reorient the data using the orientations obtained from the IMU, simply the inverse orientation is obtained then applied to each sample of the data.

There has, to date, been no communication in the research literature concerning the impact of using this method of reorientation on activity classification success-rates.

The following research questions are posed in this section:

1. **What is the impact of random monitor rotations on activity classification success rates?**

   In this question, we are interested in evaluating the impact transient monitor orientations relative to the body-location on which the monitors are mounted on, on the activity classification success-rates obtained.

   To study this impact, activity classification success-rates obtained from data that has had random rotations applied to each window are compared to those obtained from data that has had no random rotations applied to each window. Random rotations applied to each window represent an extreme case of the monitor having a different orientation relative to the body-location it is mounted on at different periods of the subject's data gathering session.

2. **What is the difference in success-rates obtained from non-reoriented data from those obtained from reoriented data?**

   In this question, we are interested in evaluating the impact of accelerometer-based reorientation of data that has had random rotations applied to each window on activity classification success-rates.

To study the impact, the success-rates obtained from data that has been reoriented are compared to success-rates obtained from data that has not been reoriented. Both sets of data had random rotations applied to each data window. Only impact of the accelerometer-based method is evaluated since this is the published and commonly referenced method.

3. **What is the difference in the success-rates obtained from data reorientated using the accelerations-based method versus success-rates obtained from data reorientated using the orientation-based method**

   In this question, we are interested in evaluating the difference in the impact of reorientation using the accelerometer-based method compared to the impact of reorientation using the orientation-based method on activity classification success-rates.

   To study the impact, the success-rates obtained from data that has been reoriented using the accelerometer based method are compared to success-rates obtained from data that has been reoriented using the orientation based method. In addition, the impact of the two methods on both data that has had random rotations applied, and data that has not had random rotations applied, is studied.

Methodology

In order to analyse the impact of random rotations and data reorientation on the activity classification success-rates; random rotations were conditionally applied to each window after segmenting the data using sliding windows; this was followed by feature-extraction using either Bao and Intille's feature-set, Kwapisz et al.'s feature-set or Henpraserttae et al.'s feature-set; finally feature-vectors were extracted and classified. For Henpraserttae's feature-set, either the original acceleration-based method proposed by Henpraserttae et al. was used to reorient the data to world coordinates, or the orientation-based method proposed in this section was used. The process of computing result sets to answer the research questions is hence as follows:

For each monitor, all three *sources* were used. A sliding window of 10 seconds with 50% overlap was used.

Window lengths of 10 seconds were used because 10 seconds was observed to be the maximum window length used in activity recognition literature, having only been used by Kwapisz et al. (2011) and Patel et al. (2009). In the analysis of activity recognition literature performed by Lockhart and Weiss (Lockhart & Weiss, 2014), window lengths reported to have been used in activity recognition literature were observed to have a median of 3 seconds and the maximum window length they observed was 10 seconds.

A 50% window overlap was used because it was observed that 50% window overlaps are common within the literature review having been used by Bao and Intille (2004); Figo et al. (2010); He et al. (2008); Krishnan and Panchanathan (2008); Kunze et al. (2005); Preece, Goulermas, Kenney, and Howard (2009); Ravi et al. (2005); Shoaib et al. (2014) and Sun et al. (2010). However, other window overlaps also exist in the literature review including: no overlap (Kwapisz et al., 2011), 20% (Reiss, 2014), 25% overlap (Henpraserttae et al., 2011), 33% overlap (Lester et al., 2005).

Optionally, random rotations were applied into each window. The random rotations are uniformly distributed over all possible rotation axes and rotation magnitudes in the range $[0°, 360°)$. The random rotations were added to the data used to answer all the research questions except for one of the data sets used to answer the first research question (the dataset without random rotations).

In addition, the data was optionally reoriented using either the accelerometer-based method explained by Henpraserttae et al. or by applying the inverse of the orientation data to reorient the data to global coordinates. The accelerometer-based method was applied on one of the data sets to answer the second research question (the reoriented data), and in the third research question (the data reoriented using the accelerometer-based method). The orientation-based method was used in the second dataset used in the third research question (the data reoriented using the orientation-based method).

The data was downsampled to frequencies ranging from 112Hz to 128Hz in intervals of 1Hz. This is because, as observed in section 4.2, changes in downsampling frequency impact the success-rate. Low pass filtering was performed prior to downsampling.

Features were then extracted as explained in the respective feature-set's paper (Bao and Intille (2004) and Kwapisz et al. (2011), refer to section 3.5 for more details). A Hamming window was applied to each window before extracting the frequency-domain features in Bao and Intille's feature set. Bao and Intille's feature-set and Kwapisz et al.'s feature-set were used to answer the first question because these feature-sets were not proposed with orientation independence in mind, hence are ideal for studying the impact of transient monitor orientations relative to the body-location mounted on.

The algorithm proposed by Henpraserttae et al.'s paper was used to answer the second and third research question. This is because Henpraserttae et al.'s algorithm is designed with orientation independence in mind and hence is ideal for studying the impact of data reorientation on activity classification success-rates. Henpraserttae et al.'s algorithm was altered so as to use either the original published acceleration-based data reorientation method, the orientation-based data reorientation method proposed in this section, or no reorientation.

The reorientation method as explained by Henpraserttae et al. is elaborated as algorithm 10.

The orientation-based method proposed in this section makes use of the orientations obtained from the IMU and hence is simple because these orientations already represent the orientation of the monitor relative to global coordinates (down from

Algorithm 10 Reorient the window W to global coordinates using accelerations A.

procedure AccelReorient(W,A)
  $G \leftarrow mean(A)$ ▷ Compute the gravity vector from the accelerations
  for all $t \leftarrow [1, 2, ..., |A|]$ **do**
    $A'_t = A_t - A_t \cdot G$ ▷ Obtain horizontal components of the accelerations
  end for
  $P_1, P_2 \leftarrow PCA(A')$ ▷ Obtain the first principal component (anteroposterior axis) and second principal component (mediolateral axis) of the horizontal accelerations.
  if $A_1 \cdot P_1 < 0$ **then**
    $P_1 \leftarrow -P_1$ ▷ Make sure the anteroposterior axis always results in the first acceleration being positive.
  end if
  for all $t \leftarrow [1, 2, ..., |W|]$ **do** ▷ Compute the projection of W onto axes P_1,P_2 and G

$$R_t = (W_t \cdot P_1)\begin{bmatrix}1\\0\\0\end{bmatrix} + (W_t \cdot P_2)\begin{bmatrix}0\\1\\0\end{bmatrix} + (W_t \cdot G)\begin{bmatrix}0\\0\\1\end{bmatrix}$$

  end for
  return R
end procedure

gravity, magnetic north from the compass and east from the cross-product of down and magnetic north). For example, the orientation of a monitor that is placed on a level flat surface facing directly north would read an orientation that is 0° pitch from down, 0° bearing (from north) and 0° roll. To reorient the data gathered back to global coordinates, it is necessary to obtain the inverse of the orientation (i.e. similar orientation in magnitude but in the opposite direction), then apply this to the data gathered. The orientation-based reorientation method is elaborated as algorithm 11.

Finally, classification was performed using the J48 decision tree from the WEKA toolkit. The J48 decision tree is an implementation of the C4.5 algorithm (Hall et al., 2009). The C4.5 decision tree was found to perform best by Bao and Intille (Bao & Intille, 2004), and second best by Kwapisz et al. (Kwapisz et al., 2011). Henpraserttae et al. did not mention about any testing performed with multiple

Algorithm 11 Reorient the window W to global coordinates using orientations O.

```
procedure AccelReorient(W,O)
    for all t ← [1, 2, ..., |W|] do
        Q ← quaternion representing the monitor orientation at time t
        Q ← inverse of Q                      ▷ Obtain inverse of the monitor's orientation
        P ← conjugate of Q
        R_t ← Q W_t P                         ▷ Rotate vector W_t
    end for
    return R
end procedure
```

classifiers, but our own testing found the J48 decision tree to perform better than KNN (with K=3) that was used by Henpraserttae et al.

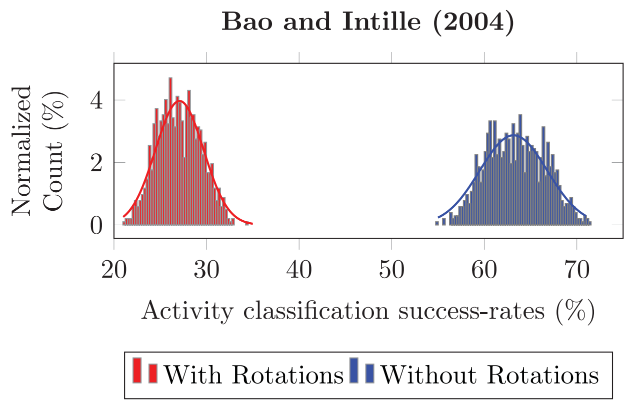
Experiments were performed using 10-fold cross-validation. For each monitor, the classifier was trained with a subset of the data obtained from that monitor and tested with the rest of the data obtained from that monitor. This is in line with the model of a phone where only one sensor is available, hence only one body-location can be monitored at any time.

Results

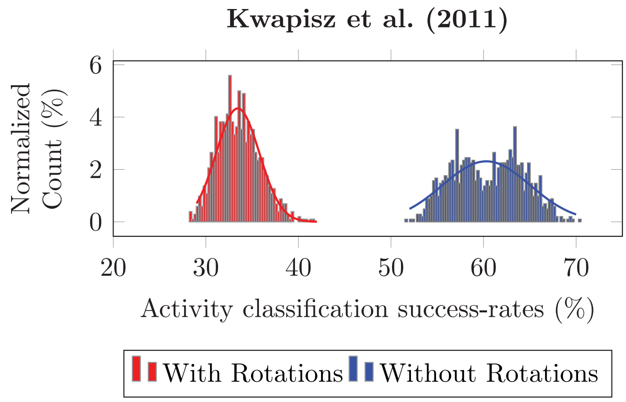
What is the impact of random monitor rotations on activity classification success rates?

Figure 7.14 shows histograms of activity classification success-rates obtained from feature-vectors extracted from data with random rotations applied to each window and data with activity classification success-rates obtained from feature-vectors extracted from data with no random rotations applied to each window. Both results obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown in the figure.

Figure 7.15 shows histograms of differences in activity classification success-rates obtained from feature-vectors extracted from data with random rotations applied

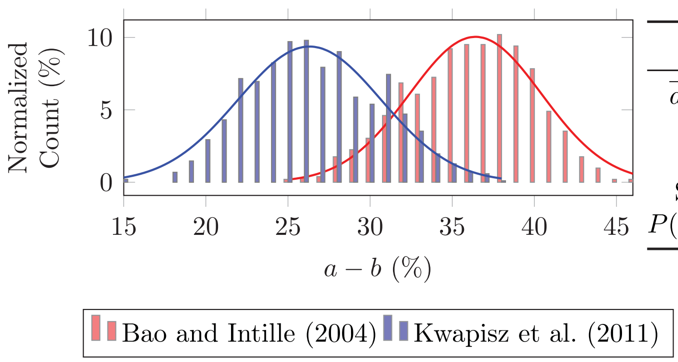


	Without Rotations	With Rotations
$\overline{x}$	63.2	27.1
σ	2.27	1.92
R^2	0.8075	0.8886
SSE	12.9	10.8



	Without Rotations	With Rotations
$\overline{x}$	60.3	33.5
σ	2.59	1.82
R^2	0.6364	0.9129
SSE	19.6	10.9

Figure 7.14: Histograms of activity classification success-rates obtained from feature-vectors extracted from data with random rotations applied to each window and data with no random rotations applied to each window. Both results obtained using Bao and Intille's feature-set (upper) and Kwapisz et al.'s feature-set (lower) are shown.



	Bao	***Kwapisz***
$\overline{a-b}$	36.41	26.31
σ	2.3853	2.4840
R^2	0.9513	0.9303
SSE	13.68	16.56
$P(a < b)$	~ 0	~ 0

Figure 7.15: Histograms of differences in activity classification success-rates obtained from feature-vectors extracted from data with random rotations applied to each window (b), from those obtained from feature-vectors extracted from data with no random rotations applied to each window (a). Differences obtained using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown.

to each window, from feature-vectors extracted from data with no random rotations applied to windows. Differences obtained using both Bao and Intille's feature-set and Kwapisz et al.'s feature-set are shown on the histograms.

From figure 7.15, it can be observed that random rotations applied to windows reduce success-rates obtained by Bao and Intille's feature-set and Kwapisz et al.'s feature-set by an average of 36.41% and 26.31% respectively. In addition, it should be noted that, based on the fitted distributions, the likelihood of a fold of data not being impacted by random rotations (i.e. a testing fold with random rotations applied resulting a higher or similar success-rate as a similar fold without random rotations applied) is almost zero.

Paired two-sample two-tailed t-tests with $\alpha = 0.05$ between success-rates obtained from feature-vectors extracted from data with no random rotations applied to windows, and success-rates obtained from feature-vectors extracted from data with random rotations applied to each window, confirmed that the two sets of success-rates came from independent random samples of normal distributions of unequal means for both success-rates obtained using Bao and Intille's feature-set and Kwapisz et al.'s feature-set.

The results show that the random rotations applied to windows had a statistically significant impact on the success-rates obtained. In addition, an average 26.31% and 36.41% decrease in success-rates can be considered large. Hence, the results show not only a statistically significant but also a considerably large impact on success-rate.

In addition, a paired two-sample two-tailed t-test with $\alpha = 0.05$ between the differences in success-rates obtained with and without random rotations, obtained using Bao and Intille's feature-set, and those obtained using Kwapisz et al.'s feature-set, confirmed the hypothesis that the two sets of differences came from independent random samples of normal distributions with unequal means.

Hence, it can be concluded that the impact of random rotations applied to windows varies depending on the feature-set used. For the two feature-sets studied, the impact was observed to be higher on success-rates obtained using Bao and Intille's feature-set than on success-rates obtained using Kwapisz et al.'s feature-set.

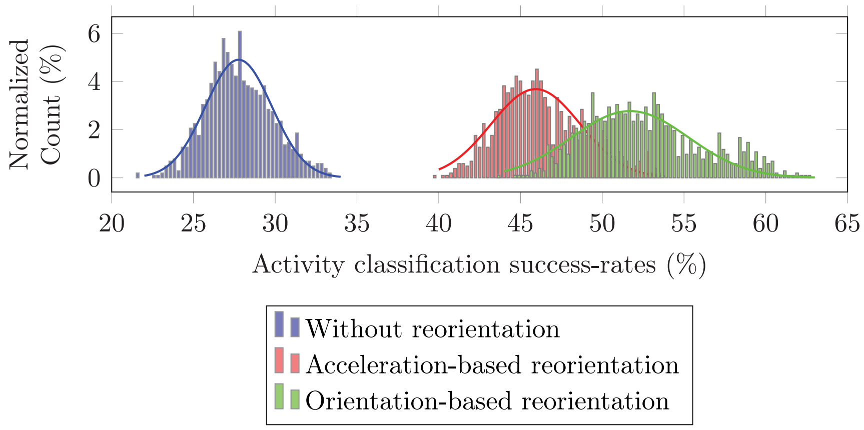
What is the difference in success-rates obtained from non-reoriented data from those obtained from reoriented data?

Figure 7.16 shows histograms of activity classification success-rates obtained from feature-vectors extracted using Henpraserttae et al.'s feature-set from data with random rotations applied to each window. Three sets of activity classification success-rates are shown: one set was obtained from data any reorientation performed, another from data with acceleration-based reorientation performed, and with orientation-based reorientation performed.

Figure 7.17 shows a histogram of differences in activity classification success-rates obtained from feature-vectors extracted from reoriented data, from feature-vectors extracted from data with no reorientation performed. Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to each window.

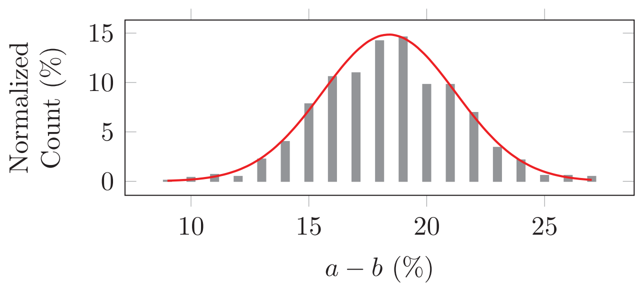
Similarly, figure 7.18 shows a histogram of relative differences of the activity classification success-rates obtained from feature-vectors extracted from reoriented data, from the activity classification success-rates obtained from feature-vectors extracted from data with no reorientation performed. The same result sets as in figure 7.17 were used.

From figure 7.15, it can be observed that reorientation of data that had random rotations applied to each window results in an average of 18.39% increase in success-rates. From figure 7.18, it can be observed that this is equivalent to an average relative increase of 39.93%.



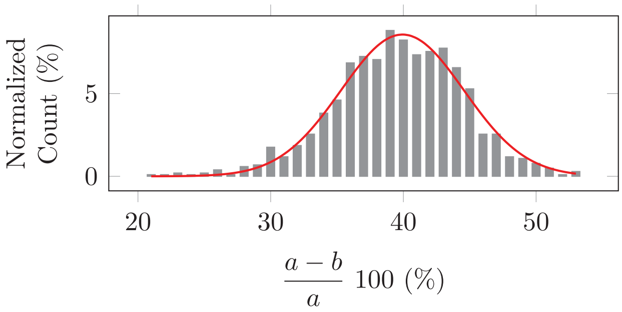
	Without reorientation	**Acceleration-based reorientation**	**Orientation-based reorientation**
$\overline{x}$	27.8	45.9	51.7
σ	1.69	1.96	2.25
R^2	0.9262	0.8254	0.7918
SSE	9.8	16.7	15.6

Figure 7.16: Histograms of activity classification success-rates obtained from feature-vectors extracted using Henpraserttae et al.'s feature-set from data with random rotations applied to each window. Three sets of results are shown: one obtained without any reorientation performed, with acceleration-based reorientation performed, and with orientation-based reorientation performed.



	Value
$\overline{a-b}$	18.39
σ	1.9972
R^2	0.9877
SSE	5.68
$P(a<b)$	~ 0

Figure 7.17: Histogram of differences in activity classification success-rates obtained from feature-vectors extracted from reoriented data (b), from those obtained from feature-vectors extracted from data with no reorientation performed (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to each window.



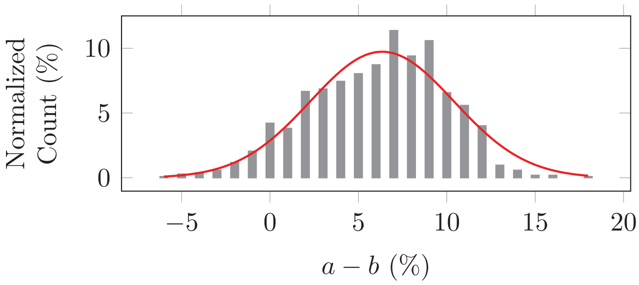
	Value
$\overline{\left(\frac{a-b}{a}\,100\right)}$	39.93
σ	2.5596
R^2	0.9744
SSE	7.7129
$P(\frac{a-b}{a} < 0)$	~ 0

Figure 7.18: Histogram of relative differences of activity classification success-rates obtained from feature-vectors extracted from reoriented data (b), from those obtained from feature-vectors extracted from data with no reorientation performed (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to each window.

In addition, it should be noted that, based on the fitted distributions in both figure 7.15 and figure 7.18, the likelihood of a fold of data that has had random rotations applied to each window, not getting impacted by performing reorientation on it (i.e. a testing fold with random rotations applied but not reoriented, resulting a higher or similar success-rate as a similar fold with random rotations applied and reoriented) is almost zero.

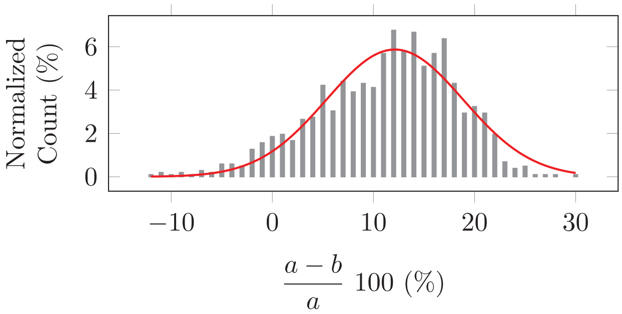
Paired two-sample two-tailed t-tests with $\alpha = 0.05$ between the success-rates obtained from feature-vectors extracted from reoriented data that had had random rotations applied to each window, and the success-rates obtained from data that that had had random rotations applied to each window but had not been reoriented, confirmed that the two sets of success-rates came from independent random samples of normal distributions of unequal means.

Hence, from the results obtained, it can be concluded that reorientation applied on data that had had random rotations applied to each window resulted in a significant increase in success-rates.



	Value
$\overline{a-b}$	6.33
σ	2.3966
R^2	0.9381
SSE	20.50
$P(a < b)$	0.0041

Figure 7.19: Histogram of differences in activity classification success-rates obtained from feature-vectors extracted from data reoriented using the original proposed accelerometer-based method (b), from those obtained from feature-vectors extracted from data reoriented using orientations (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to each window.

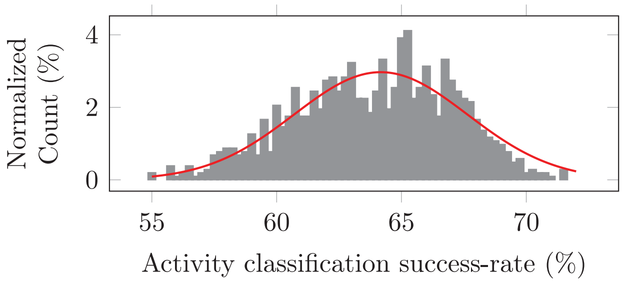


	Value
$\overline{\left(\frac{a-b}{a}100\right)}$	12.11
σ	3.0910
R^2	0.9161
SSE	16.38
$P(\frac{a-b}{a} < 0)$	0.00004

Figure 7.20: Histogram of relative differences of activity classification success-rates obtained from feature-vectors extracted from data reoriented using the original proposed accelerometer-based method (b), from those obtained from feature-vectors extracted from data reoriented using orientations (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to each window.

What is the difference in the success-rates obtained from data reorientated using the accelerations-based method from success-rates obtained from data reorientated using the orientation-based method?

Figure 7.19 shows a histogram of differences in activity classification success-rates obtained from feature-vectors extracted from data reoriented using the original proposed accelerometer-based method, from feature-vectors extracted from data reoriented using orientations. Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that had random rotations applied to



	Value
$\overline{x}$	64.2
σ	2.21
R^2	0.8198
SSE	0.46

Figure 7.21: Histogram of activity classification success-rates obtained from feature-vectors extracted using Henpraserttae et al's feature-set from data with no random rotations applied to each window and no reorientation performed.

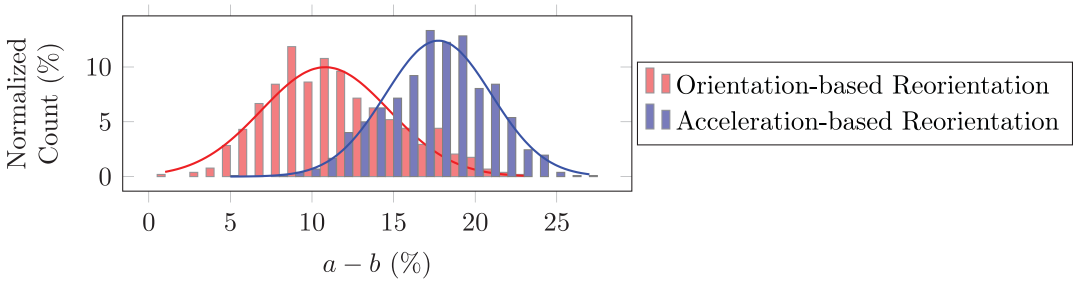
each window.

Similarly, figure 7.20 shows a histogram of relative differences of activity classification success-rates obtained from feature-vectors extracted from data reoriented using the original proposed accelerometer-based method, from feature-vectors extracted from data reoriented using orientations. The same result sets as in figure 7.19 were used.

From figure 7.19, it can be observed that data reorientation using orientations result in higher success-rates than data reorientation using accelerations for the majority of the test folds. The mean difference between success-rates obtained from feature-vectors extracted from data reoriented using the original proposed accelerometer-based method, from success-rates obtained from feature-vectors extracted from data reoriented using orientations, is 6.33%. This is equivalent to an average 12.11% relative increase in success-rates.

From both figure 7.19 and figure 7.20, it can be observed that a small proportion of the test folds are likely to result in a higher success-rate by using accelerations to reorient the data than by using orientations to reorient the data. However, this proportion is very small compared to the proportion of test folds that result in higher success-rates by using orientations to reorient the data instead of accelerations.

Figure 7.21 shows a histogram of activity classification success-rates obtained from feature-vectors extracted using Henpraserttae et al's feature-set. The data



	Orientation-based Reorientation	**Acceleration-based Reorientation**
$\overline{a-b}$	10.81	17.74
σ	2.3526	2.1399
R^2	0.9129	0.9540
SSE	24.84	18.93
$P(a < b)$	~ 0	~ 0

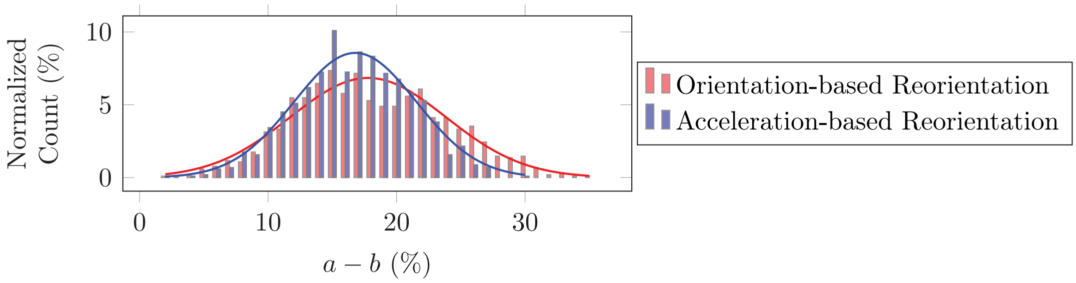
Figure 7.22: Histogram of the differences in activity classification success-rates obtained from feature-vectors extracted from data that had random rotations applied to each window then reoriented using accelerations or orientations (b), from those obtained from feature-vectors extracted from data that did not have any random rotations applied (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set.

used to compute figure 7.21 had no random rotations applied to each window. In addition, no data reorientation was performed.

Figure 7.22 shows a histogram of the difference between success-rates obtained from feature-vectors extracted from data without any random rotations applied to windows, to those obtained from feature-vectors extracted from data that has had random rotations applied each window then reorientated using the orientation-based method and the acceleration-based method.

From the figure, it can be observed that data that has had no random rotations applied results in higher success-rates than either result set that has had random rotations applied to each window then reoriented using either the orientation-based reorientation or acceleration-based reorientation.

Similarly, figure 7.23 shows a histogram of the differences in activity classification success-rates obtained from feature-vectors extracted from data that has been reoriented using accelerations or orientations, from feature-vectors extracted from



	Orientation-based Reorientation	**Acceleration-based Reorientation**
$\overline{a-b}$	17.78	16.78
σ	2.9270	2.5978
R^2	0.9207	0.9817
SSE	14.98	4.58
$P(a < b)$	~ 0	~ 0

Figure 7.23: Histogram of the differences in activity classification success-rates obtained from feature-vectors extracted from data that has been reoriented using accelerations or orientations (b), from those obtained from feature-vectors extracted from data that has not been reoriented (a). Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that did not have any random rotations applied.

data that has not been reoriented. Feature-vectors of both data sets were extracted using Henpraserttae et al.'s feature-set from data that did not have any random rotations applied.

Similar to figure 7.22, it can be observed that data that has had no random rotations applied and no reorientation performed, results in higher success-rates than data that has had no random rotations applied but has had reorientation performed by either the orientation-based reorientation or acceleration-based reorientation.

From the fitted distributions showed in figure 7.22 and figure 7.23, it can be observed that it is unlikely that a test fold that has had either acceleration-based reorientation or orientation-based reorientation performed, results in a higher success-rate than a test fold that has had no random rotations and no reorientation performed.

Hence, it can be concluded that the highest success-rates are likely to be obtained when neither random rotations are applied to data nor reorientation performed on the data. Reorienting the data using either the accelerations-based method or the orientation-based method results in lower success-rates. However, as it has been shown, the success-rates obtained from data that has been reoriented are higher than those obtained from data that has had random rotations but not been reoriented.

Discussion

First, the impact of random rotations on activity classification success-rates was studied. Success-rates were obtained using the two feature-sets studied throughout the thesis (Bao and Intille (2004) and Kwapisz et al. (2011)). In addition, success-rates were also obtained using the same feature-sets using data that had random rotations applied to each window. The differences in the two sets of success-rates were then obtained.

Application of random rotations onto the data resulted in a decrease from 63.2% to 27.1% activity classification success-rates for feature-vectors extracted using Bao and Intille's feature-set and 60.3% to 35.5% activity classification success-rates for feature-vectors extracted using Kwapisz et al's feature-set. Activity classification success-rates obtained using Henpraserttae et al.'s feature-set decreased from 64.2% to 27.8%. When accelerometer-based reorientation was applied to the data before extracting feature-vectors using Henpraserttae et al.'s feature-set, the activity classification success-rate increased to 45.9%. Orientation-based reorientation resulted in an activity classification success-rate increased of 51.7%.

In addition, it was also noted that the likelihood of a test fold, computed from data with random rotations applied to each window, having a higher success-rate than a similar test fold, but computed from data without random rotations applied to the windows, is almost zero.

It is hypothesised that the cause of the large decrease in success-rates observed when random rotations are applied to each window, is that the random rotation cancel out any similarities within data of activities where the differences in the data captured by the sensor is largely due to the orientation of the sensor.

However, if random rotations are applied to each window, each window is transformed using a random rotation hence resulting in data that has different statistical features to the original data. Hence, the 10 windows result in 10 feature-vector clusters. The classifier then requires characteristics of 10 clusters to identify the walking up stairs activity.

Next, the impact of reorienting data that has had random rotations applied to each window was studied. To do this, the activity classification algorithm proposed by Henpraserttae et al., that reorients data using the gravity vector extracted from the accelerations and the PCA of the horizontal accelerations, was altered to allow both reorienting and non-reorienting of the data. The success-rates obtained while reorienting were then compared to those obtained without reorienting the data.

While the application of random rotations has the consequence that any classifier cannot exploit orientation information in the data to determine the current activity, data reorientation attempts to undo these effects by transforming the data captured from the local sensor coordinates to global coordinates. For example, vectors that were originally directed upwards would be encoded relative to the x, y and z-axes of the sensor. Since the orientation of the sensor changes from window to window, this would result in different recorded x, y and z values of the vector depending on the orientation of the sensor. By making use of the gravity vector to compute the rotation necessary to reorient the data captured by the sensor to global coordinates, then applying this rotation to reorient the data, all the vectors that were originally directed upwards get mapped to similar coordinates independent of the orientation of the sensor.

Next, the difference in success-rates obtained using accelerometer-based reorientation from those obtained using orientation-based reorientation is studied. This was performed by altering Henpraserttae et al. algorithm such that the same feature-set is used but reorientation is performed using orientations instead of accelerations. The success-rates obtained were then compared to those obtained by using the original algorithm that uses accelerations to reorient the data.

The results showed that a mean increase in success-rate of 6.33% is obtained while using orientations to reorient than by using accelerations to reorient the data. This is equivalent to an average relative increase in success-rates of 12.11%. A small proportion of the testing folds were observed to have a larger success-rate while using acceleration-based reorientation than by using orientation-based success-rates. However, the proportion was observed to be very small compared to the proportion of the data that results in higher success-rates while using orientations to perform data reorientation.

It is hypothesised that the better success-rates obtained using orientations to perform data reorientation than by using accelerations to perform data reorientation are due to the orientation-based reorientation resulting in a more accurate reorientation than when accelerations are used. The acceleration-based method of reorientation proposed by Henpraserttae et al. makes use of the gravity vector that is encoded in accelerations to reorient the data to global coordinates. However, this method estimates the gravity vector using the mean of the accelerations recorded in the whole window. Hence, if the orientation of the sensor relative to gravity in one part of the window is different from that of another part of the window, the resulting gravity vector is the mean obtained from those different sensor orientations. Which actually does not match with any of the original orientations but would lie somewhere between them depending on the proportion of the time the sensor was in one orientation compared to the other. Orientations, however, are derived from integrating the rotational velocities obtained from the gyroscope then including the gravity vector computed from the accelerations to correct the eventual drift. This results in an orientation estimation for each sample in the window that is not impacted by the window computation.

When the success-rates obtained from feature-vectors extracted from data that had been reoriented were compared to success-rates obtained from feature-vectors where no random rotations were applied and no reorientation was applied, it was observed that the reorientation of data that either had had random rotations applied or not, resulted in lower success-rates than when the data had no random rotations and no reorientation performed.

Hence, it can then be concluded that the highest success-rates are likely to be obtained when the orientation of the monitor is fixed relative to the body-location (as is the case when wearable sensors are used). Changing the orientation of the monitor relative to the body-location it is mounted on results in a decrease in success-rate, even when either acceleration-based reorientation or orientation-based reorientation is performed.

However, if the orientation of the monitor relative to the body-location is transient (such may be the case when a mobile phone is used), then reorientation is likely to result in higher success-rates than would otherwise have been obtained. Hence, reorientation ameliorates the decrease in success-rates due to the transient nature of the orientation of the monitor relative to the body-location it is monitoring.

In addition, using orientation-based reorientation results in higher success-rates than using acceleration-based reorientation. It is hypothesised that this is due to the higher accuracy of the monitor orientations obtained using the orientation-based method compared to the monitor orientation obtained using the acceleration-based method.

The random rotations applied to the data in this section are an extreme case. This is because it is unlikely that the number of orientations possible for a phone in a pocket would surpass the uniformly distributed random rotations over the whole range of possible orientations used. The orientation of the phone is likely to be constrained by the size and shape of the pocket. Even when the size of the pocket is large, the phone is likely to settle at the bottom of the pocket and hence result in a smaller number of orientations. In addition, it is unlikely that the phone would change orientations on each window (every 5 seconds) in the day-to-day life of the phone user.

7.3 Conclusion

In this chapter, studies are performed on location independence and orientation independence in smart-phone-based activity recognition. Within smart-phone-based activity recognition, two main challenges exist: the location of the phone on the subject's body is transient (hence any solution needs to incorporate location independence) and the orientation of the phone relative to the body-location it is also transient (hence any solution needs to incorporate orientation-independence).

For location independence, a fundamental research question was asked: whether the body-location on which a monitor is mounted can be identified without knowing which activity the subject is performing. Next, the success-rates of identifying the body-location were evaluated and compared with reference to the subject's activity. In addition, an evaluation was performed to find out whether the success-rates of identifying the body-location the monitor are mounted on are dependent on the orientation of the monitor.

The results obtained showed that the body-location on which a monitor is mounted can be identified with high success-rates (86.42% and 87.58% for Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively). A break-down of the success-rates obtained showed that each body-location could be identified with high success-rates (above 80% for all body-locations and both feature-sets studied).

Next, the accuracy of identifying the body-location on which a monitor is mounted during each activity was studied. It was observed that the success-rates varied from one activity to the next, but were high for all activities (above 60% for all activities and both feature-sets used). The activities were ranked based on the likelihood of achieving a higher success-rate of identifying body-locations during one activity more than during another activity. Among the top ranked activities for both feature-sets were writing, using the PC, texting on the phone and watching TV. Among the lowest ranked activities for both feature-sets were talking on the phone, dusting and vacuuming.

Next, a study was performed on how the accuracy of identifying the body-location a monitor was mounted on, depends on the orientation of the monitor. To do this, random rotations were introduced into each data window. This simulates the monitor being in a different orientation relative to the body-location it was mounted on, in each window.

It was observed that the success-rates of identifying the body-location dropped significantly (to 45.66% and 54.85% for Bao and Intille's feature-set and Kwapisz et al.'s feature-set respectively) when random rotations were introduced into each data window processed. However, the success-rates were still significantly above chance (16.67%). Hence, it was concluded that the orientations are only partially responsible for the ability to differentiate the body-location a monitor was mounted.

For orientation independence, the impact of random monitor rotations on activity classification success-rates was first studied. This was performed by comparing the activity classification success-rates obtained with and without random rotations applied to the data.

Application of random rotations onto the data resulted in a decrease from 63.2% to 27.1% activity classification success-rates for feature-vectors extracted using Bao and Intille's feature-set and 60.3% to 35.5% activity classification success-rates for feature-vectors extracted using Kwapisz et al's feature-set. Activity classification success-rates obtained using Henpraserttae et al.'s feature-set decreased from 64.2% to 27.8%. When accelerometer-based reorientation was applied to the data before extracting feature-vectors using Henpraserttae et al.'s feature-set, the activity classification success-rate increased to 45.9%. Orientation-based reorientation resulted in an activity classification success-rate increased of 51.7%.

Hence, it was concluded that when the monitor orientation was fixed relative to the body-location is it mounted (as is the case with wearable sensors), then reorientation results in decreased success-rates. However, if the monitor's orientation relative to the body-location is it mounted is transient (as is the case with phones), reorientation results in increased success-rates. In addition, between the two reorientation methods, the orientation-based method results in slightly higher success-rates than the accelerometer-based method.

8 Conclusion

In this thesis, an analysis of activity recognition on Activities of Daily Living using on-body inertial sensors was performed. Accelerations, rotational velocities and orientations captured from monitors mounted on subjects while they performed the activities in semi-controlled environments were studied. Two feature-sets were used to evaluate several fundamental classification settings that impact activity classification accuracy using on-body inertial sensors. These include:

1. The impact of sampling frequency on the activity classification accuracy and the sampling requirements of each *source*.
2. The impact of sliding window length on the activity classification accuracy.
3. The impact of sliding window overlap on the activity classification accuracy.
4. The difference in activity classification success-rates obtained by the two selected feature-sets.
5. The difference in activity classification success-rates obtained by each individual *source* (accelerations,rotational-velocities and orientations) and the best *source* to use.
6. The difference in activity classification success-rates obtained by each individual body-location and the best body-location to use.
7. The impact of the number of body-locations monitored on the activity classification accuracy.
8. The differences in accuracy of identifying different activities and the hierarchy

of activities from the most easily identified to the most difficult to identify.

9. The most easily confusable activities and the hierarchy of confusability of activities.
10. A comparison of remove-one-subject cross-validation results to 10-fold cross-validation results that indicates the the impact of inter-subject variation.
11. Identification of the body-location on which a monitor is mounted independent of the user's activity.
12. The impact of sensor orientation on the identification of the body-location on which a monitor is mounted.
13. The impact of random monitor rotations on activity classification success-rates.
14. The impact of data reorientation from local coordinates to global coordinates on activity classification success-rates.
15. The difference in impact of different data reorientation methods on activity classification success-rates.

In order to answer the research questions connected to the classification settings mentioned, data was gathered of 22 Activities of Daily Living in a semi-controlled environments. The data included accelerations, rotational velocities and orientations sampled at 128Hz. In addition, data from a six body-locations (three for two of the activities) was captured for each subject. The sensors were synchronised to $\leq 10\mu s$ sample timing difference. For 20 of the 22 activities, six body-locations were gathered while for 2 of the activities only three body-locations were gathered. Where necessary, analysis was performed using six body-locations for the 20 activities, and using three body-locations for 22 activities. Each activity had a minimum of 18 subjects. Walking and running had the highest number of subjects: 21.

Two feature-sets were selected so as to strengthen the findings. Data processing and classification was performed independently using each feature-set and patterns in the resulting activity classification or location classification success-rates were analysed to deduce the impact of various parameters. The C45 decision tree classifier was used for both feature-set because it was observed to perform well for both feature-sets. The analysis of the resulting success-rates not only takes into consideration which parameter results in higher average success-rates but how consistently the parameter results in higher average success-rates.

In the following sections, the findings from each major analysis will be summarised. In addition, the implications of the findings and future research work will be discussed.

8.1 Analyses performed and findings

Characteristics of sliding windows in data segmentation

The analysis results show that the longer the sliding window length the higher the activity classification success-rates obtained. Hence, a sliding window length should be picked as the longest length that is practical for the activity recognition system. For the window overlap, the larger the window overlap the higher the activity classification success-rates obtained using 10-fold cross-validation. However, window overlaps were found not to impact success-rates obtained using remove-one-subject cross-validation.

The sliding window length

It was observed that there is a linear relationship between the success-rate obtained and the logarithmic window length used to extract the feature-vectors for both feature-sets studied. This is a diminishing returns relationship between the success-rate obtained and the window length used to extract the feature-vectors.

This implies that a larger window length always results in either similar or larger success-rates. However, it also implies that larger increases in window lengths are required to obtain the same increments in success-rates as found at at lower window lengths.

This relationship is explained as a result of sampling error. As the window length increases the sampling error in the statistical model obtained from the sample decreases. However, as the window length increases, the likelihood of encountering new information that is not already contained in the window grows lower with increase of window length.

However, practical constraints also have to be considered while picking a window length. A larger window length not only results in more data to be processed (and hence more computational resources used and more battery power consumed), but also results in a lag between the time when the action was performed and the time it is recognised. The absence of lag might be critical for real-time systems.

In addition, longer window lengths are more likely to include data from multiple activities during transitions from one activity to another. This can decrease the system's accuracy during transitions and possibly make detecting transitions from one activity to another more difficult.

The sliding window overlap

It was observed that there is a decreasing linear relationship between the success-rate obtained and the logarithmic shift (or lag) between windows when 10-fold cross-validation is performed. In addition, the variance is observed to decrease with increasing window overlap. However, this relationship does not exist when remove-one-subject cross-validation is performed.

This relationship is explained as the result of information shared by windows. When the windows highly overlap, a significant portion of the information contained in one window is also contained in the adjacent windows. When, these adjacent windows are used to train the classifier while the current window is used to test the classifier (as is the case with 10-fold cross-validation), higher success-rates are obtained.

Any increase in success-rate obtained specificly due to shared information in adjacent windows being in both the testing set and the training set can be thought of as increments due to over fitting the data and hence false increments in success-rates. This is because, this exact common information shared between training set and testing set is unlikely to occur naturally due to the usage of the system by an end user.

However, as noticed in section 4.4, the mean success-rate obtained while performing 10-fold cross-validation with no window overlap is higher than the success-rate obtained while performing remove-one-subject cross-validation. In addition, the standard deviation of success-rates obtained while performing 10-fold cross-validation with no window overlap is lower than the standard deviation of success-rates obtained while performing remove-one-subject cross-validation. This implies that while window overlaps result in an increase in success-rates, the presence of the subject's data in both the training set and the testing set also results in increased success-rates. The presence of the subject's data in both the training set and the testing set (as well as data from other subjects) is the hybrid model as discussed in section 2.6 that is trained using both the current system user's data and additional data from other subjects.

Whether to use accelerations, rotational velocities, orientations or all three

The average MES frequencies (see section 4.2 for definition) over all studied activities of accelerations, rotational velocities and orientations are 6Hz, 10Hz and 2Hz respectively for Bao and Intille's feature-set and 8Hz, 9Hz and 3Hz respectively for Kwapiz et al.'s feature-set.

Although MES frequencies as high as 63Hz were observed, the differences between the minimum sampling requirements of accelerations, rotational velocities and orientations in order to achieve the highest success-rates were found to be statistically significant but too small to impact any activity recognition system (the largest difference in sampling frequency was 1.43Hz). However, of the three *sources*, orientations were noted to require the lowest sampling frequencies. In terms of success-rates achievable by any of the three *sources*, the analysis shows that all three *sources* should be used if possible. If not possible, then accelerations should be used since success-rates obtained by accelerations are only marginally lower than those of the three *sources* combined.

It was observed that the highest success-rates obtained from feature-vectors extracted from rotational velocities are obtained at marginally but significantly higher sampling frequency than either feature-vectors extracted from accelerations or orientations for both feature-sets studied. It was also observed that highest success-rates obtained from feature-vectors extracted from orientations are obtained at marginally but significantly lower sampling frequency than either feature-vectors extracted from accelerations or rotational velocities for both feature-sets studied. Hence, of the three *sources* studied, rotational velocities require the highest sampling frequency while orientations require the least sampling frequency.

The impact of a higher sampling frequency is in terms of larger power consumption and larger use of computational resources. Higher sampling frequencies result in more data captured which in turn result in more data processed. However, the differences in the sampling frequency requirements of the three *sources* is only marginal (averaging 1.43Hz at most between any two *sources*). It is unlikely that such a marginal difference would impact battery consumption, even for the current resource-constrained wearable sensor systems.

Hence, the selection of which *source* to use need not depend on the sampling frequency requirements but should depend on the ability of the features derived from the *sources* to discriminate between activities.

Feature-vectors extracted from accelerations were observed to result in higher activity classification success-rates than feature-vectors extracted from either rotational velocities or orientations. Feature-vectors extracted from orientations were found to result in higher success-rates than feature-vectors extracted from rotational velocities.

When success-rates obtained from feature-vectors extracted from each individual *source* were compared to success-rates obtained from feature-vectors extracted from the three *sources* combined, it was observed that feature-vectors extracted from accelerations resulted in success-rates that were only marginally lower than those from the three *sources* combined.

A possible explanation as to why accelerations and orientations perform much better than rotational velocities is the presence of orientational information in both accelerations and orientations that is absent from rotational velocities. This additional orientational information can be used to distinguish between activities where some of the motions captured by the monitors are similar but performed in different sensor orientations. For example, the thigh monitor while sitting and standing captures the same stationary motion, but differs in orientation.

The results obtained imply that in situations where attaining the highest success-rates is important, if all three *sources* can be obtained and processed, then combining all three is likely to result in a higher success-rate than any one individual *source*. However, if this is not possible, then accelerations should be prefered to orientations, which in turn should be preferred to rotational velocities.

The number of monitors to mount and the best body-locations to mount them on

For a single monitor

The analysis shows that, for a single monitor, either one of the wrists or the thigh should be used.

Feature-vectors extracted from data captured from the wrists and thigh result in higher success-rates than data captured from the chest, dominant upper arm or ankle. This is likely due to a large number of the activities studied having arm motions and the ability of the thigh sensor to distinguish sitting activities from standing activities.

This implies that if one location was to be selected to recognise Activities of Daily Living, then this body-location should be either the thigh or one of the wrists. The best body-location can vary depending on the feature-set. For Bao and Intille's feature-set, the differences in the success-rates obtained from data captured from the three body-locations were statistically insignificant. However, for Kwapisz et al.'s feature-set, data captured from the dominant wrist resulted in higher success-rates than data captured from the non-dominant wrist, which in turn resulted in higher success-rates than data captured from the thigh monitor.

For multiple monitors

The analysis shows that, activity classification success-rates obtained from more than three monitors are only marginally higher than those obtained by three monitors. The three monitors should be mounted on the wrists and thigh to achieve the highest activity recognition success-rates.

It was observed that there is a diminishing returns relationship between the number of monitors used and the activity classification success-rates. This implies that with the right sensor placement, the first few monitors contribute the most to the activity classification success-rate. Each additional monitor contributes less information than the preceeding monitor.

The best body-location to place the first three monitors is observed to be either one of the wrists or the thigh, depending on the feature-set used. The increments in success-rates past the third monitor are observed to be marginal.

Hence, to accurately identify Activities of Daily Living, it is advisable to mount three monitors on the subject: one on each wrist, and one on the thigh.

The most easily recognised activities and the most difficult to recognise

The analysis shows that running, using the PC, watching TV, writing and walking on a flat surface are the most easily recognised activities (i.e. have the highest success-rates). The activities with the lowest success-rates are washing hands and washing vegetables.

When the activity classification success-rates of identifying each activity were analysed, it was observed that activity classified with the highest success-rate is running, followed by using the PC, watching TV, writing and walking on a flat surface. The activities recognised with the lowest success-rate are washing vegetables and washing hands; followed by dusting and sweeping; then vacuuming, ironing, washing dishes and folding clothes.

In addition, it was observed that three distinct groups of activities existed: a group with high success-rates and low variance in success-rates; a group with medium success-rates and medium levels of variance in success-rates; and a group with low success-rates and high variance in the success-rates.

The first group was observed to include running, walking, walking up stairs, walking down stairs, watching TV, writing, using a PC and texting on the phone.

The second group was observed to include grating, dicing, stiring, talking on the phone, brushing teeth, peeling vegetables, vacuuming, ironing, washing dishes, folding clothes, dusting and sweeping. The third group was observed to include washing hands and washing vegetables only.

The differences in mean success-rates and standard deviation of success-rates likely result in the ability of the feature-sets to discriminate between the activities. The most highly recognised activities are also the activities most consistently recognised. This implies that there are particular characteristics of the data of these activities that is captured by the feature-sets resulting in being accurately and consistently recognised. The lowest recognised activities are observed to be very similar to each other: washing hands and washing vegetables. It is possible that these activities can not be differentiated by data captured from the sensors used and additional information like whether or not the person is washing in the bathroom or the kitchen, could be used to better distinguish between them.

Highly and mutually confusable activities

The greatest mutual confusion exists between sweeping and vacuuming; followed by folding clothes and ironing; then washing hands and washing dishes.

When the mutual confusion rates between activities were analysed, it was observed that the greatest mutual confusion existed between sweeping and vacuuming. It is likely that of all the activities studied, sweeping and vacuuming have the most similar gross motor motions to each other. Additional sensor data, either via the use of RFID tags and readers or the use of a microphone, could be used to recognise the presence of the vacuum cleaner and hence distinguish between the two activities. Of the two approaches, the microphone is a more prevalent technology than RFID tags and readers since each smart phone contains one. However, this approach would possibly not be effective in a noisy environment.

The next highest mutual confusion is observed to be between folding clothes and ironing, and between washing hands and washing dishes, depending on the result set.

Between ironing and folding clothes, it was observed that ironing has motions like folding and flipping motions that are prevalent in folding clothes. In order to reduce the mutual confusion between ironing and folding, it is proposed that classification could focus on identifying motion patterns in ironing that are not present in folding (e.g. moving the iron back and forth).

Similarly, between washing hands and washing dishes, it is likely that some motions exist in one activity but not the other. For example, scrubbing motions might only exist in washing dishes but not in washing hands. Hence, it is necessary to explore the motions of the two activities and isolate motion patterns present in one activity but not present in the other.

The presence of these patterns implies that the activity currently being performed is likely one activity and not the other. For example, identification of forward and backward wrist motions could favour ironing instead of folding, while the identification of scrubbing motions could favour washing dishes instead of washing hands. One possible method of performing this is by using time-series shapelets. Time-series shapelets refer to sections in one class of time-series that are only present in the time-series of one class but not in the time-series of another class. The shapelets algorithm proposed by Ye and Keogh (2009) offers a way to identify the shapelets, and the logical shapelets proposed by Mueen, Keogh, and Young (2011) offers a way to construct a decision-tree-like binary classifier based on time-series shapelets.

The impact of inter-subject variation

10-fold cross-validation results are higher and vary less than remove-one-subject cross-validation results due to inter-subject variation. However, the relative performance of activities (to other activities) is similar between the two cross-validation methods.

When activity classification success-rates obtained from remove-one-subject cross-validation were compared to activity classification success-rates obtained from 10-fold cross-validation, it was observed that: centroids of success-rates obtained from 10-fold cross-validation were higher than those of success-rates obtained from remove-one-subject cross-validation; and dispersion of success-rates obtained from 10-fold cross-validation was lower than that of success-rates obtained from remove-one-subject cross-validation.

However, the median success-rates of individual activities obtained from 10-fold cross-validation were observed to be highly correlated to those of individual activities obtained from remove-one-subject cross-validation.

If we combine these findings and those observed in section 4.4 (the analysis of the impact of window overlap), we can observe that the presence of the subject's data in both the training set and the testing set results in success-rates that are higher and less dispersed than when the subject's data is only present in the testing set but not in the training set. This represents the hybrid model discussed in section 2.6 that is trained using both the current system user's data and additional data from other subjects. Results obtained from remove-one-subject cross-validation are representative of results obtained by the impersonal model discussed in section 2.6. The results show that the hybrid system results in higher and more consistent success-rates than the impersonal model (for the number of subjects studied) but the relative accuracy of identifying activities using the two models is similar.

The difference in success-rates between remove-one-subject cross-validation and 10-fold cross-validation (with no window overlaps) is due to inter-subject variability. The motion patterns of each subject are slightly different from the motion patterns of other subjects. The differences are smaller with simple activities but grow larger when complex activities (activities with many different motions) are considered.

This is because complex activities not only include differences in motion patterns but also the order and sequence of motions taken. It is possible that with a higher number of subject, higher and more consistent success-rates would also be obtained by using the impersonal model (remove-one-subject cross-validation).

Identifying where a monitor is mounted on the subject's body

The location on the subject's body on which a monitor is mounted on can be identified independent from the subject's activity. This is in part due to differences in the orientation of the monitors mounted in different body-locations while performing activities, and in part due to differences in the motions of different body-locations while performing activities.

By combining the data obtained from each of the six monitors and testing to identify the body-location on which the monitor that captured a test feature-vector was on, it was observed that the body-location of the monitor that captured the test sample can be identified with high success-rates.

It is possible to explain this behaviour as the result of motion (or orientation) differences between different body-locations during an activity. Within the studied Activities of Daily Living, it was observed that the body-location could be identified accurately for each of the activities. The success-rates does vary depending on the activity, but for all activities location classification was significantly above chance. This implies that there are enough differences in the motions (or orientations) captured at different body-locations during each activity, to identify the body-location.

The most easily identified body-location was observed to be the ankle, followed by the chest or thigh; then the non-dominant wrist; dominant wrist and finally the dominant upper arm.

It was also observed that the activities that resulted in higher success-rates in identifying the body-location of the monitor were activities that were highly stationary. This indicated the likelihood of to the success-rates being due to the orientation of the monitor rather than the motions captured during that activity.

To test this, random rotations were introduced into each window before extracting feature-vectors. By introducing random rotations into each window, the data captured in each window is made to appear as if the monitor changed orientations at each window. Hence, undoing any impact of the sensor orientation being fixed during data capture. The results showed a significant decrease in success-rate, but the success-rates were still significantly above chance. Hence, it was concluded that high success-rates of identifying the body-location on which the monitor is mounted are partially due to differences in sensor orientations from one monitor to another during the activity, and partially due to characteristics of the motions present at different body-location during the activity.

As future work, it would be interesting to analyse the impact of reorienting the data from local coordinates to global coordinates on the body-location identification success-rate. It is likely that reorientation would result in success-rates that are higher than those observed when random rotations were introduced into each window, but lower than those observed when the orientation of the sensor was fixed (no random rotations introduced). However, the results would more closely resemble the achievable body-location identification success-rates when smart phones are used as sensing nodes. This is because, the orientation of phones in the carry location is not fixed and phones can be carried in one of many different locations on the subject.

The impact of random monitor rotations

Random monitor rotations result in decreased activity classification success-rates. Data reorientation (from sensor coordinates to global coordinates) can alleviate this decrease. However, success-rates obtained without any random rotations results are higher than those obtained when data reorientation is performed.

The data captured from monitors that were constantly changing orientations was simulated by applying random rotations to each window before feature-extraction. The activity classification success-rates obtained showed a significant decrease from those obtained without random rotations applied to each window.

The significant decrease in success-rates is explained as resulting from the removal of orientation information that can be used to differentiate some of the activities (e.g. sitting from standing using the orientation of a monitor mounted on the thigh). In addition, the random rotations also diversify the number of models to be learned by the classifier since effectively, rotating the captured data results in the statistical characteristics of the data captured in different axes of the sensor changing.

By using reorientation methods, the data can be reoriented from local coordinates to global coordinates. This effectively results in similar data captured while the sensor was in different orientations having similar statistical characteristics to each other. Hence, results in the data not having any information about the orientation of the sensor relative to the body-location it was mounted on, but having unified models of the motions carried out during the activity. The application of reorientation methods to the data with random rotations applied to each window resulted in a significant increase in success-rates obtained.

Two methods of data reorientation were tested: an accelerometer-based method proposed by Henpraserttae et al. (2011); Mizell (2003); Yang (2009), and a proposed IMU-based method. The accelerometer-based method uses the gravity component obtained from accelerations and the PCA of the horizontal accelerations to obtain vectors of the world orientation. The vectors are then used to rotate the data captured to world coordinates. This has the disadvantage that the world coordinate vectors are computed for the whole window, which could possibly include rotations.

The IMU-based method uses orientations obtained from the IMU to reorient each sample in the window. By continuously integrating rotational velocities obtained from the gyroscope and then incorporating the gravity vector obtained from the accelerations from the accelerometer and the direction to the magnetic north obtained from the magnetometer to offset the eventual drift due to integration, the IMU is able to maintain an accurate global orientation for each sample captured. Reorienting the data using the IMU-based method resulted in significantly higher success-rates than those obtained when the acceleration-based method is used.

However, success-rates obtained from data that had neither random rotations applied nor data reorientation performed resulted in higher success-rates than when either method of data reorientation was applied. This implies that the best activity-recognition success-rates are obtained when the monitors have a fixed orientation relative to the body-location they are mounted on and changes in orientation results in decreased success-rates.

In some situations, such as when a smart phone is used as a sensing node, the orientation of the sensor is guaranteed to change. In those cases, the use of data reorientation methods results in higher success-rates than when not used. If the monitor has an accelerometer, gyroscope and magnetometer, then combining the outputs of the three sensors to obtain an orientation results in better success-rates than when the accelerometer alone is used to reorient the data.

8.2 Final words

This thesis has looked at the impact of various configuration settings on activity classification success-rates. Clearly it is not possible to examine the interaction of all combinations of these parameter settings. For example, how do the window overlap and window lengths impact the success-rates obtained after using the acceleration-based reorientation method. Future work would include analysis of the interaction of the parameters.

In addition, the analysis of this thesis assumed that the monitors did not move freely relative to the body-location mounted on. However, in reality smart phones carried in pockets are free to move around in the pocket, and sensors worn by the subject can be loose and hence move independent to the body-location they are worn on. The additional motion increases the noise in the signal and it is unclear how tolerant the data processing methods used are to noise. An analysis with both monitors firmly attached and loosely attached (so as to allow varying degrees of free movement) to the body-locations being monitored should to be conducted to characterise the impact of the sensor's free motion relative to the body-location on activity classification success-rates.

Analysis performed in section 6.4 observed significant differences between the success-rates obtained using 10-fold cross-validation from remove-one-subject cross-validation. The difference is caused by inter-subject variability and implies that the number of subjects used in this analysis is smaller than it needs to be to fully capture the inter-subject variability that exists in on-body inertial data while performing Activities of Daily Living. Further research in the area should include more subjects so as to capture the inter-subject variability.

Analysis performed in section 4.3 has shown that longer window lengths during data segmentation results in higher activity classification success-rates. However, the analysis did not take into account the occurrence of transitions from one activity to another. Longer window lengths are more likely to encounter transitions than shorter window lengths. In addition, longer window lengths result in more data getting processed by the system and introduce lag between the occurrence of an activity and when the system recognises it. Further research should take into account these practical factors when performing segmentation and possibly use window lengths that are dynamic.

Other avenues of research in the area include the use of time-series shapelets

(Ye & Keogh, 2009) to discover segments of the data that discriminate between highly confusable activities (like washing hands, dishes and vegetables). Shapelets can be used as a data exploratory tool to identify the motions within the activity, and the body-locations at which the motions occur, that differentiate one activity from another. This is especially important for complex activities that involve many different motions.

Similar to time-series shapelet, time-series motifs (Mueen, Keogh, Cash, & Westover, 2002) are subsequences of a time-series that frequently occur within the time-series. Time-series motifs can assist in the discovery and building of an alphabet that can later be used to encode more complex activities by their building block motions. The idea of using an alphabet of motions to encode and recognise activities was proposed by Ghasemzadeh et al. (2008). However, no research has been performed in the area.

To summarise the results of the analysis performed in this thesis, the following selection of parameters are suggested for system developers interested in developing on-body activity recognition systems that use inertial sensors for recognising Activities of Daily Living:

1. Mount monitors on at least three body-locations. These should include the wrists and thigh.
2. Using accelerations only can result in success-rates that are almost as high as using accelerations, rotational velocities and orientations.
3. The highest window length that is practical should be used.
4. The hybrid model should be used. This implies that the system should have the capability to incorporate the system user's data into it's trained models. However, it means that the system does not need to be trained before usage, unlike a system that uses a personal model.
5. If Bao and Intille's feature-set is used with accelerations, the highest sampling rate necessary for the dominant wrist is 16Hz (due to grating), thigh is 13Hz

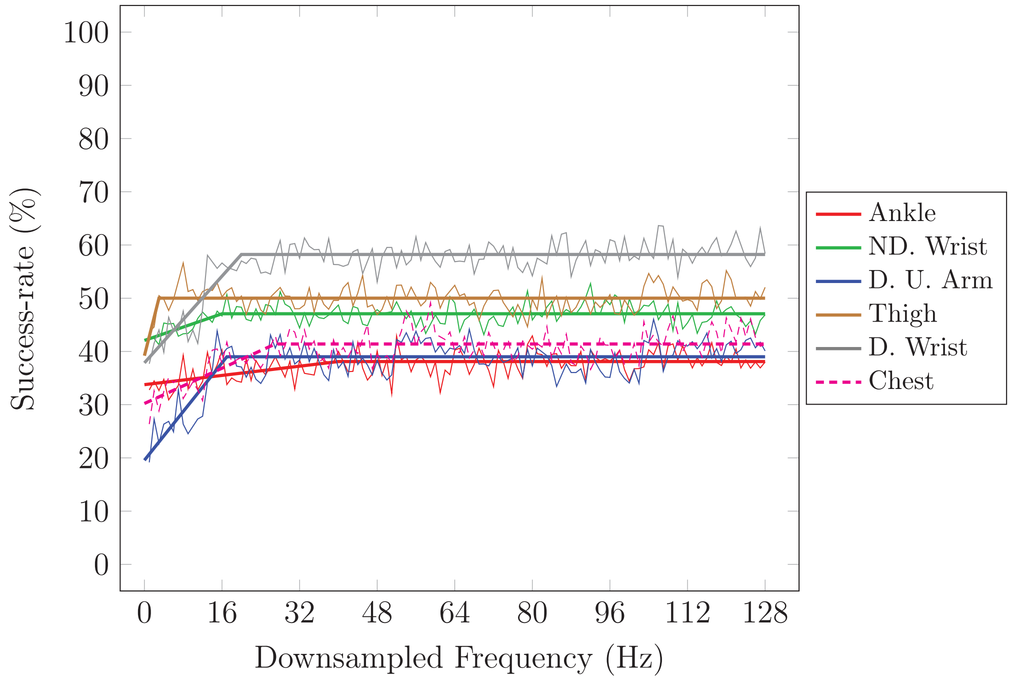
(due to grating), and non-dominant wrist is 34Hz (due to washing vegetables). If Kwapisz et al.'s feature-set is used, the highest sampling rate necessary for the dominant wrist is 63Hz (due to sweeping), thigh is 27Hz (due to vacuuming), and non-dominant wrist is 34Hz (due to grating). However, if a different feature-set is used, analysis should be performed to find the performance of the feature-set with increase in sampling frequency and the sampling frequency picked appropriately.

6. Washing hands, dishes and vegetables should either be combined, or an additional classifier (e.g. by using time-series shapelets) should be used to differentiate between the activities. In addition, the same should be performed for vacuuming, sweeping and dusting.
7. The monitors should have a fixed orientation relative to the body-locations mounted on (i.e. should be strapped to the body-location) to achieve the best success-rate.
8. In case the monitors are likely to change orientation freely, data reorientation methods should be applied to ameliorate the impact of sensor rotations on the activity classification success-rate.
9. In addition, in case the monitors are likely to change orientation freely, using an IMU to track the orientation of the sensor, results in higher success-rates than when an accelerometer only is used.
10. In case the monitors are likely to change location (e.g. when a smart phone is used as sensing node), the location of the monitor on the body can be accurately identified, even when the orientation of the monitor is not fixed relative to the body-location.

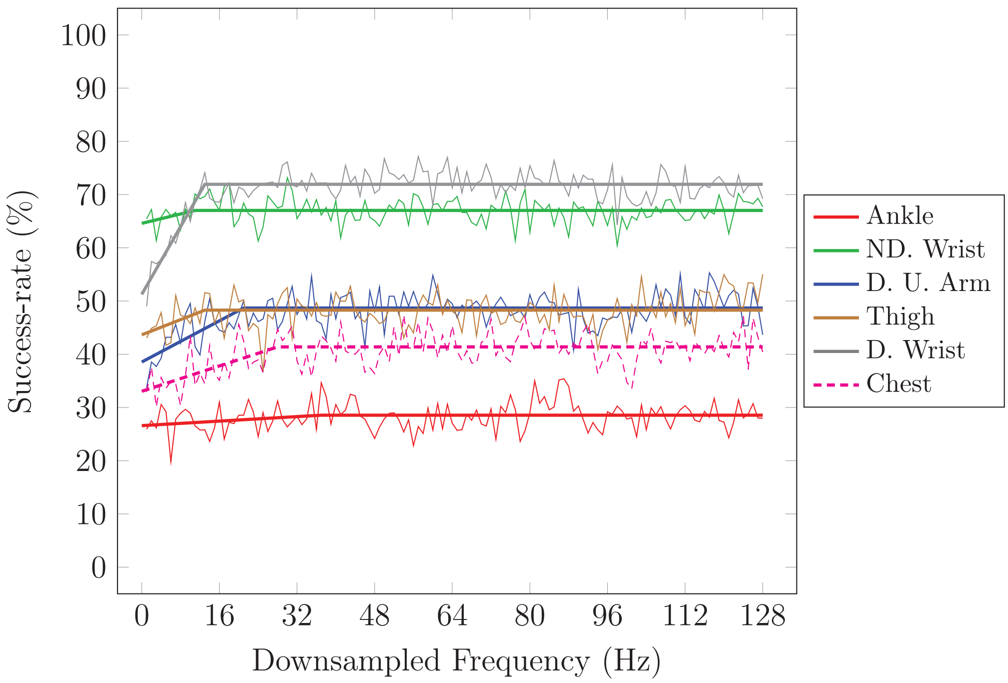
A Downsampled activity success-rate charts

Bao and Intille (2004) - Accelerometer

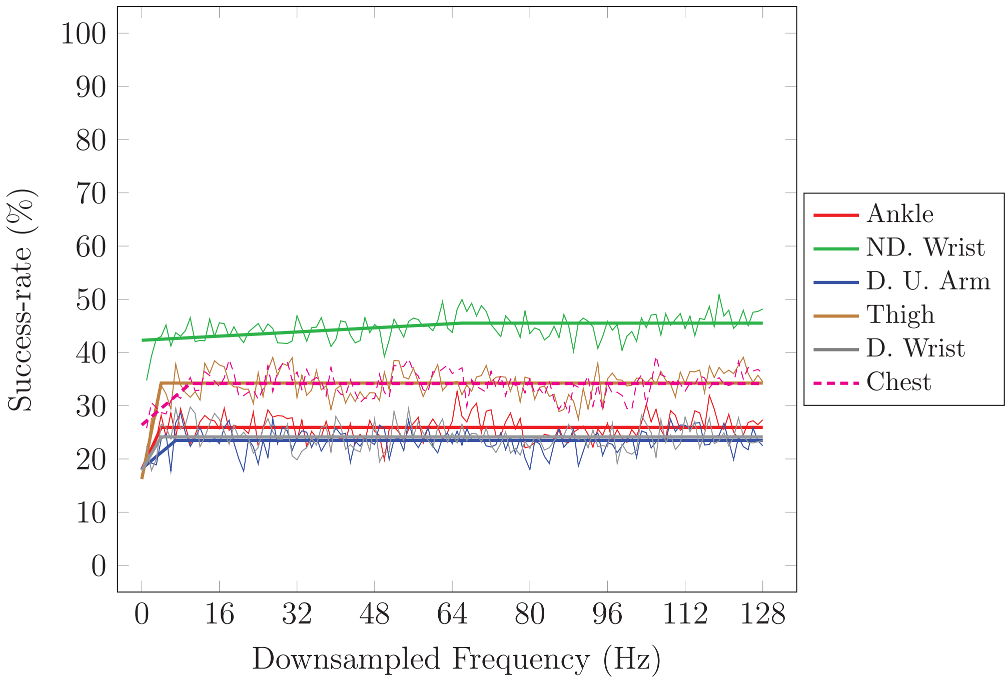
Brushing Teeth



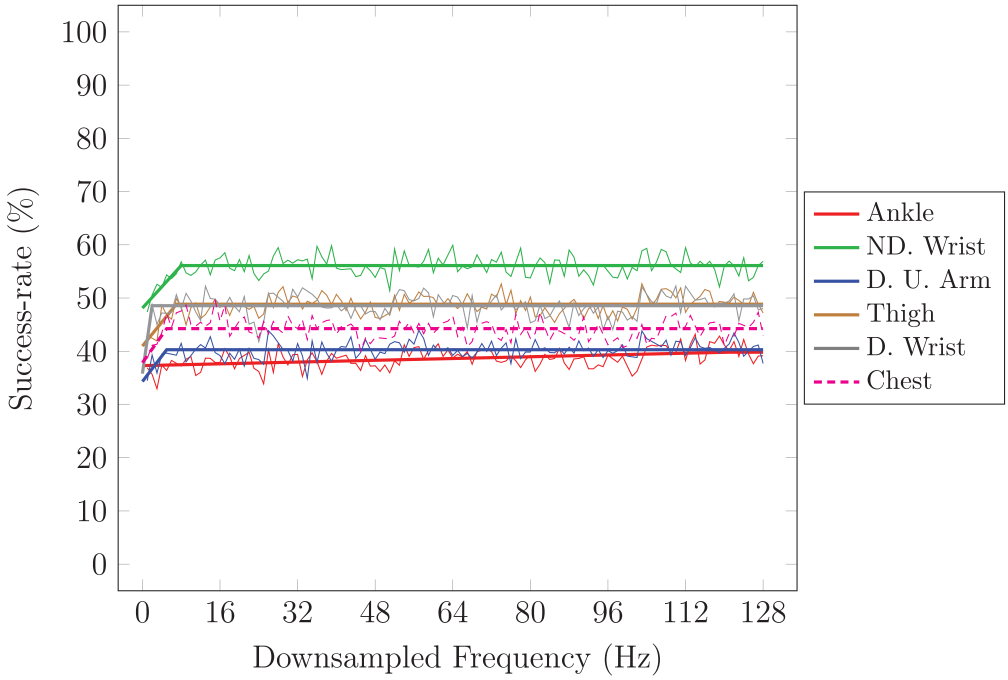
Dicing



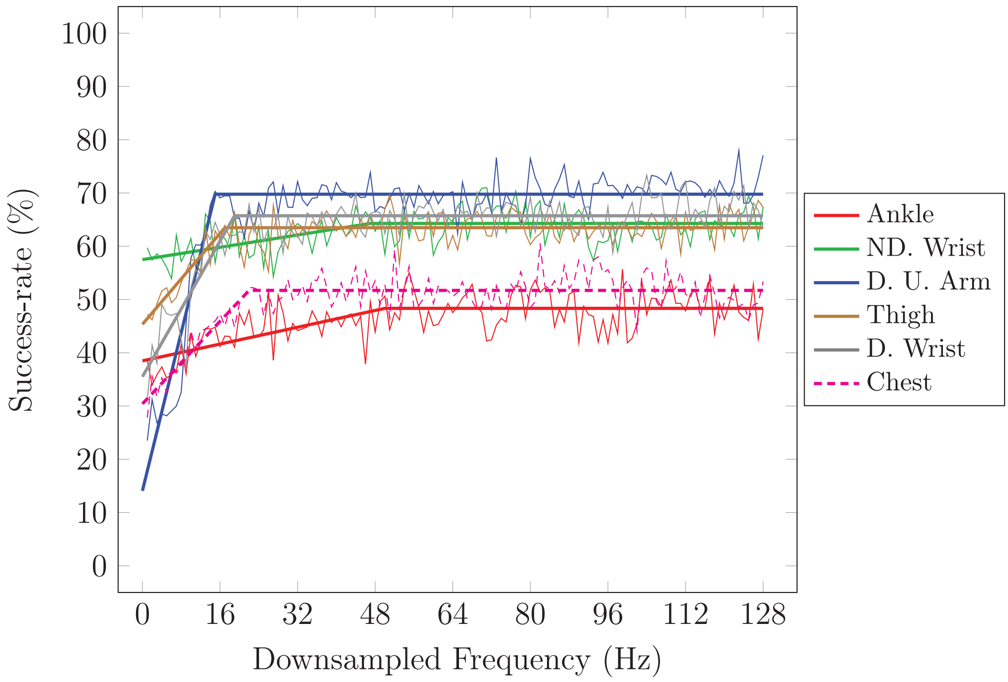
Dusting



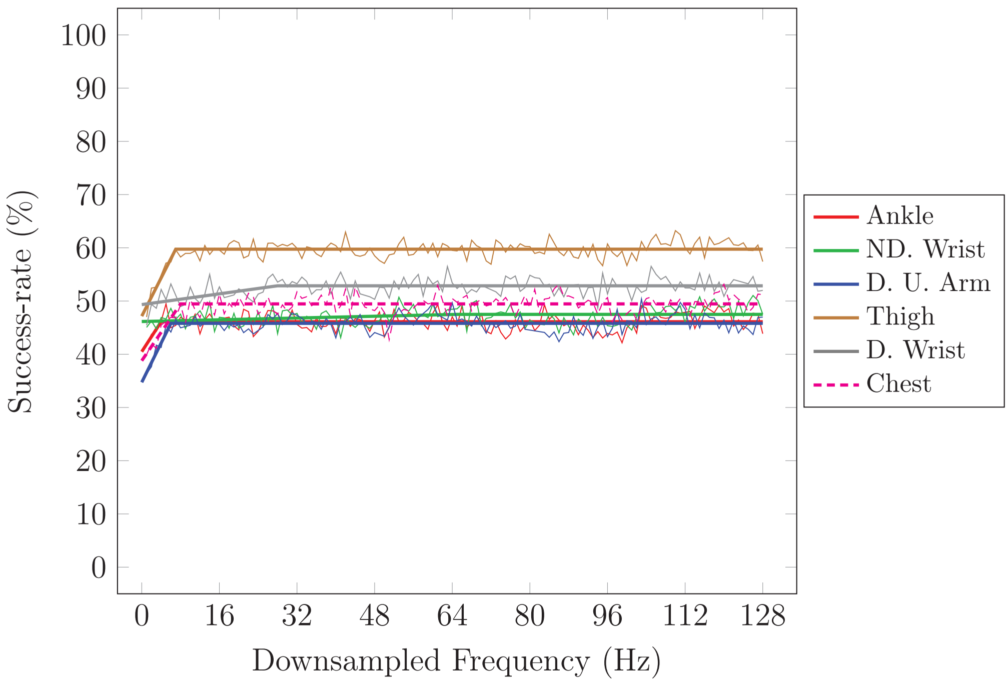
Folding Clothes



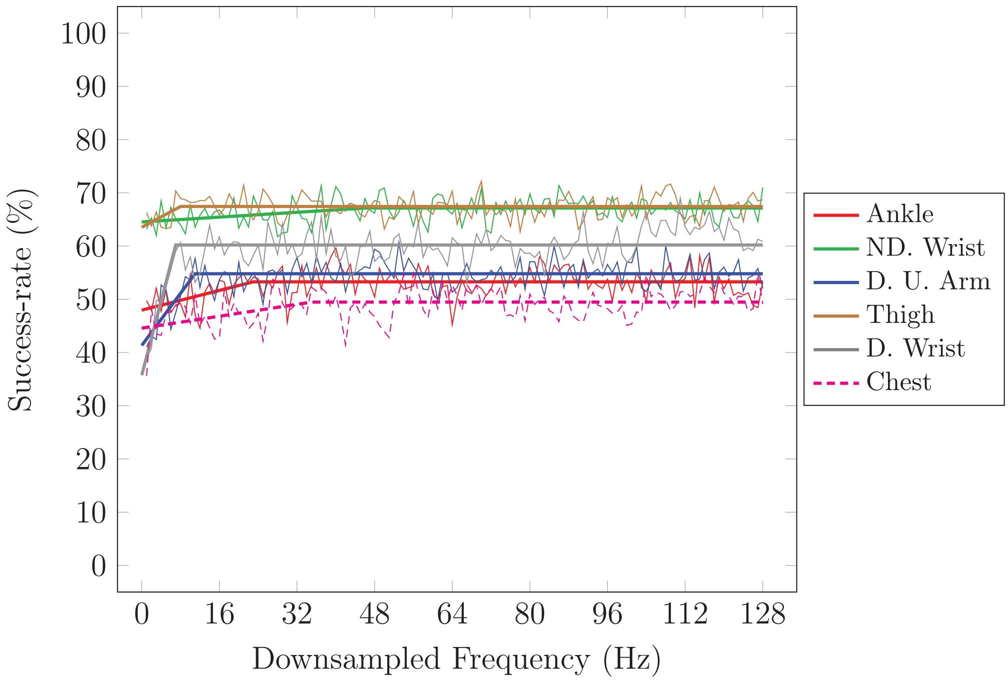
Grating



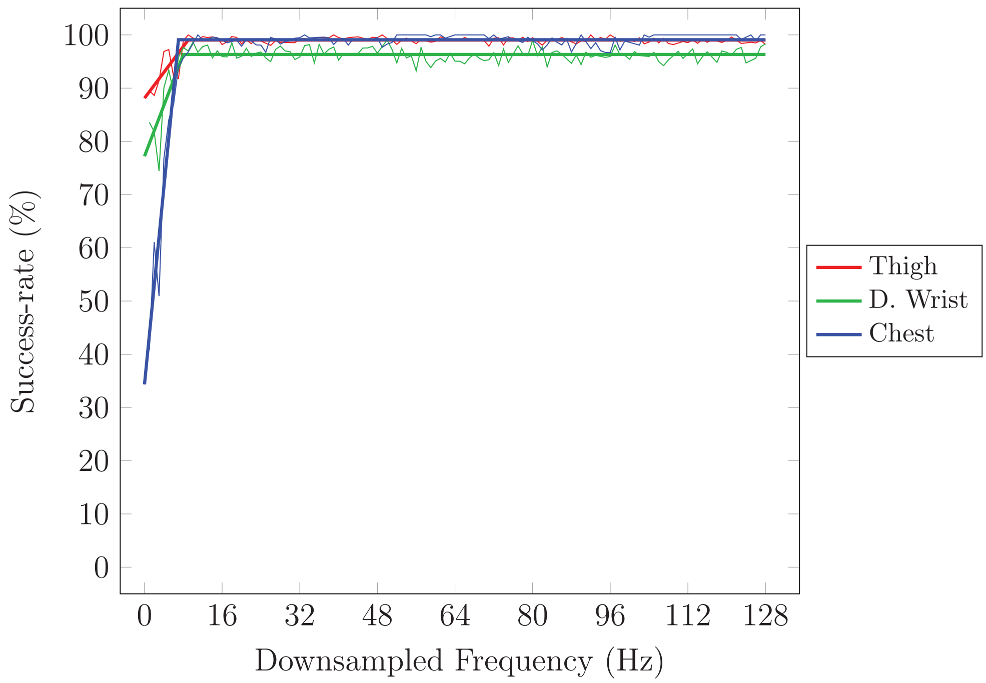
Ironing



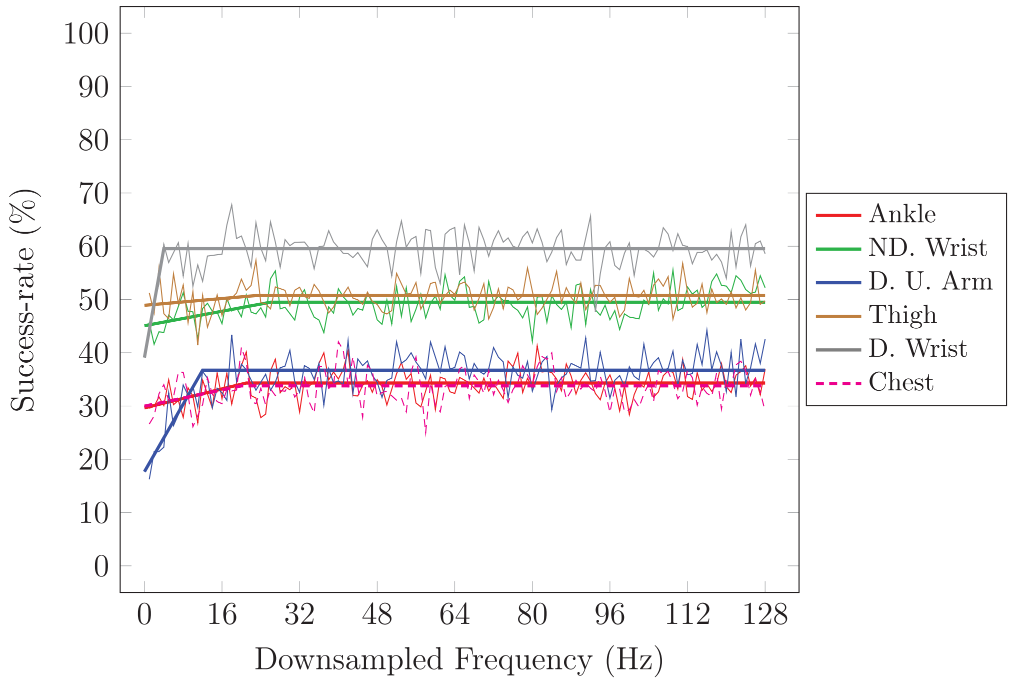
Peeling Vegetables



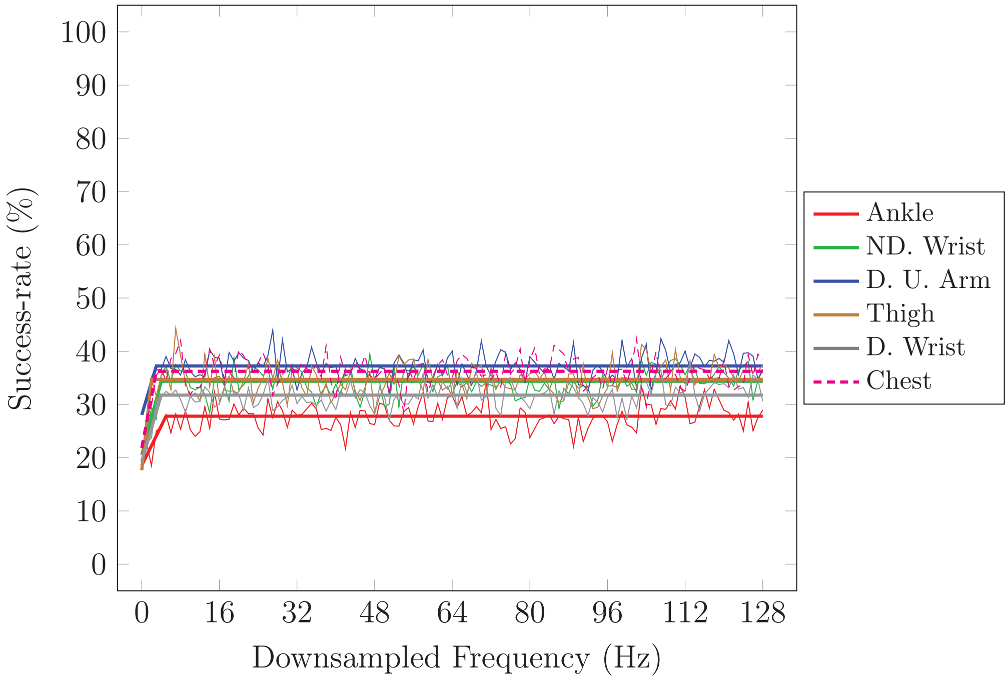
Running



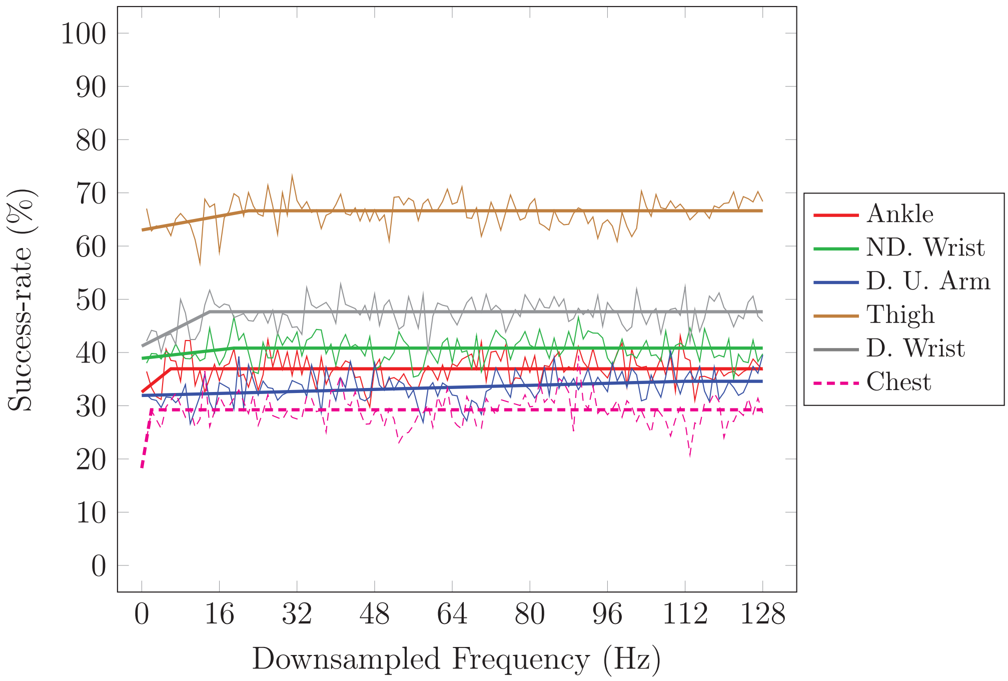
Stiring



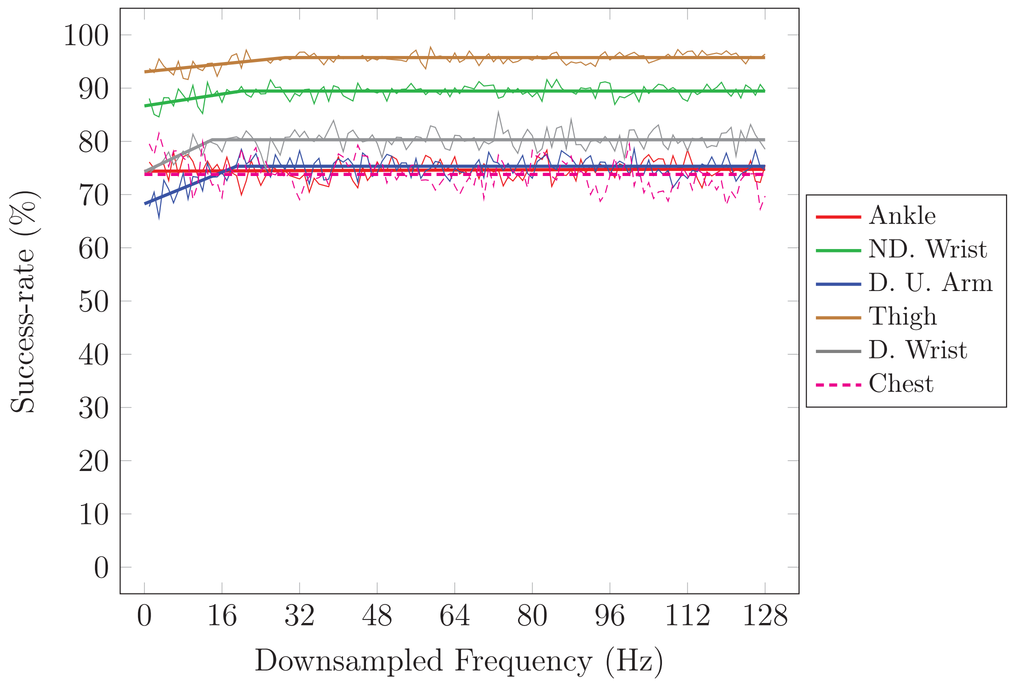
Sweeping



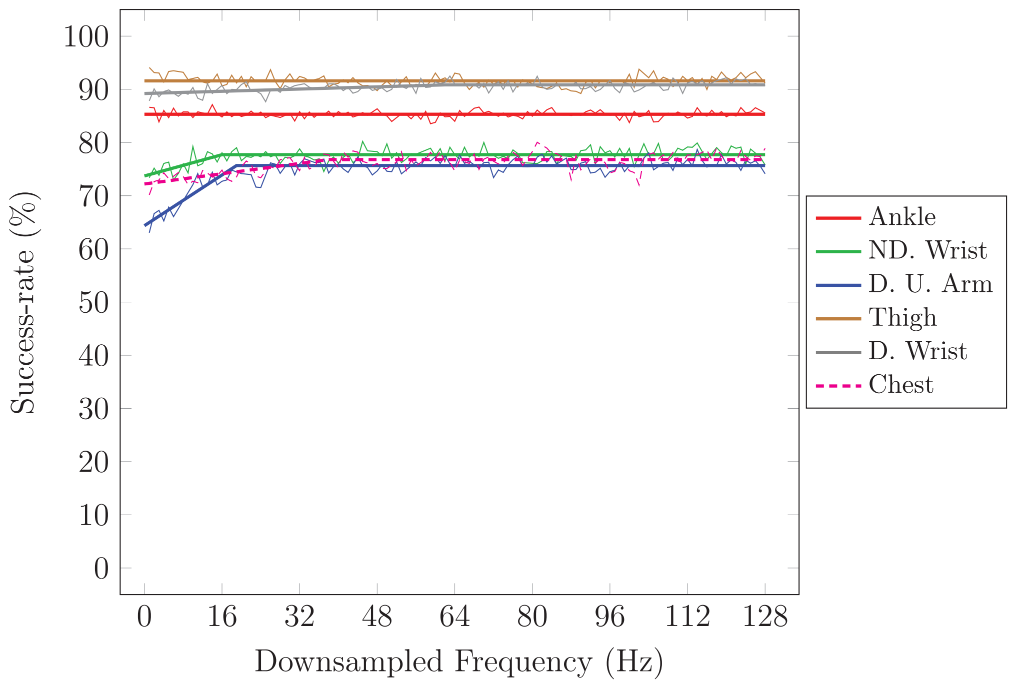
Talking on a Phone



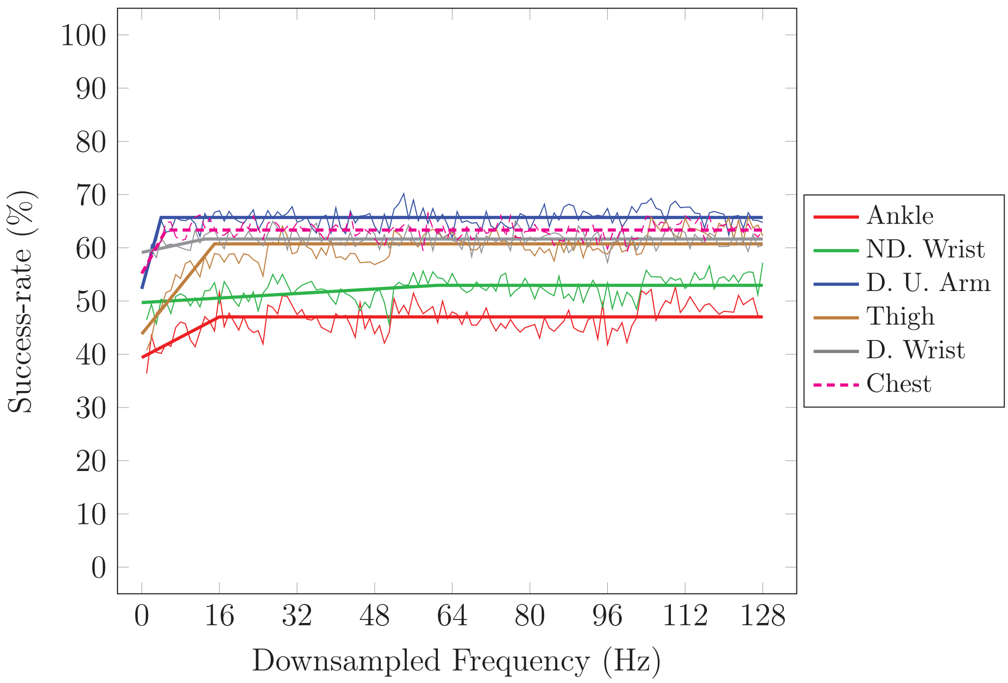
Texting on a Phone



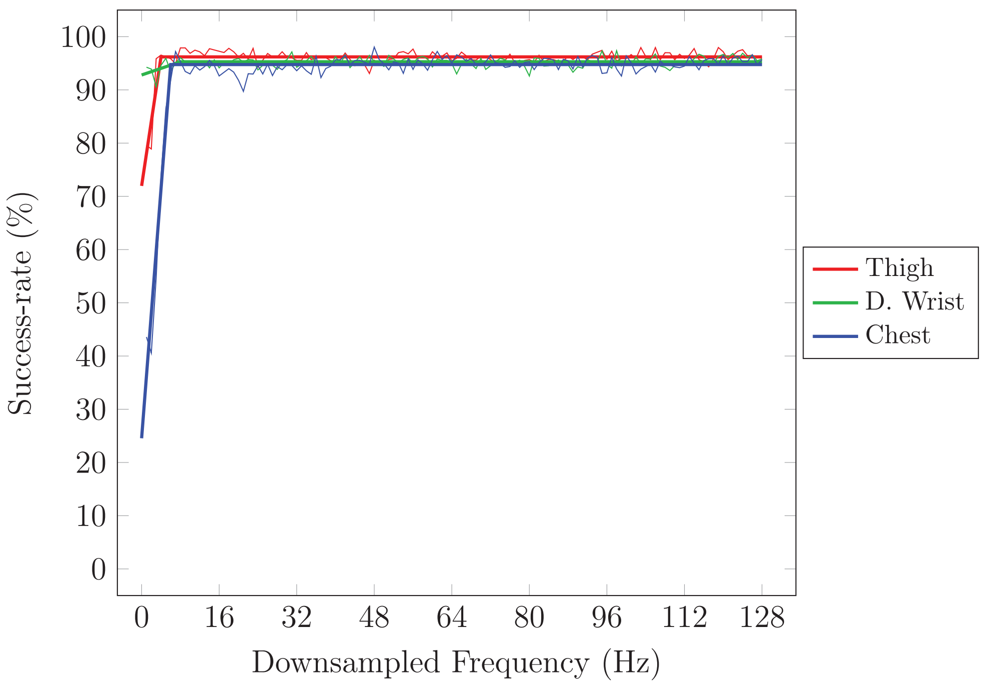
Using a PC



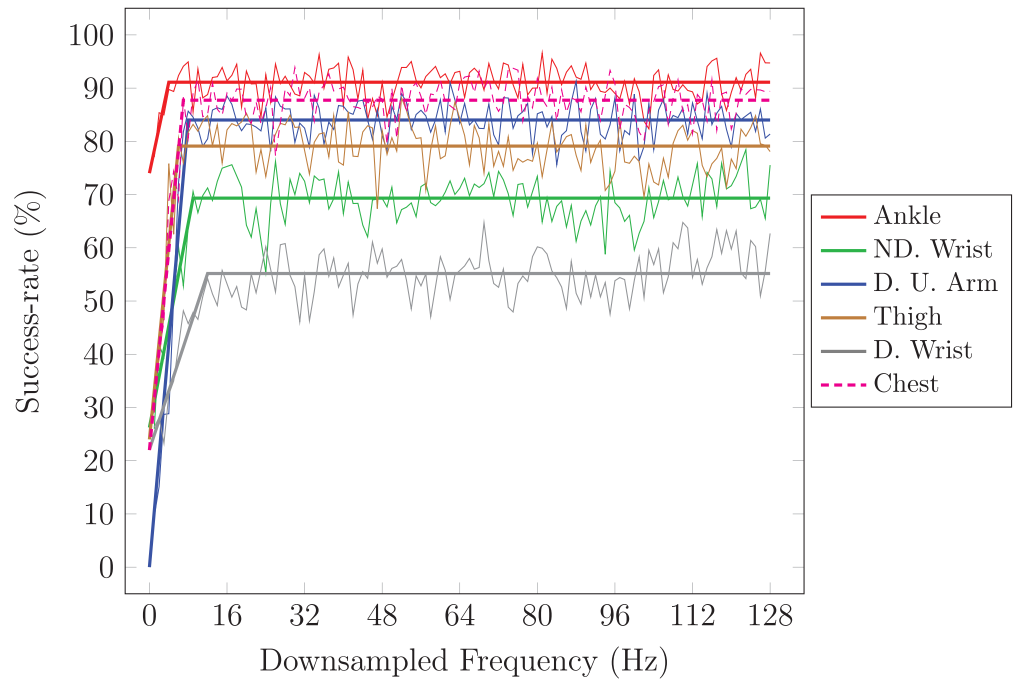
Vacuuming



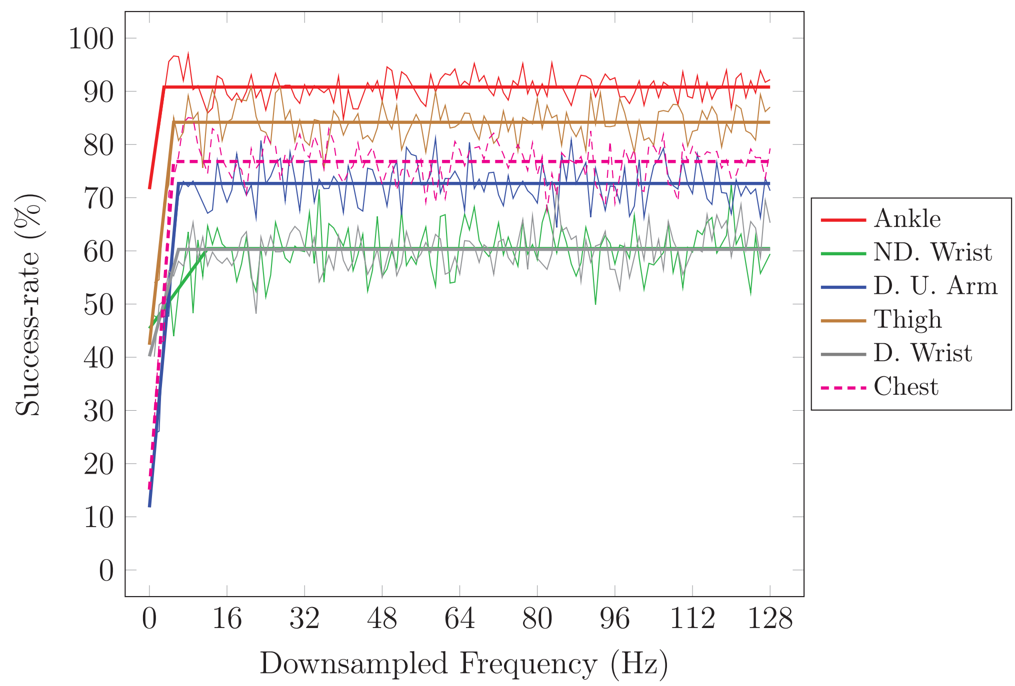
Walking (Flat Ground)



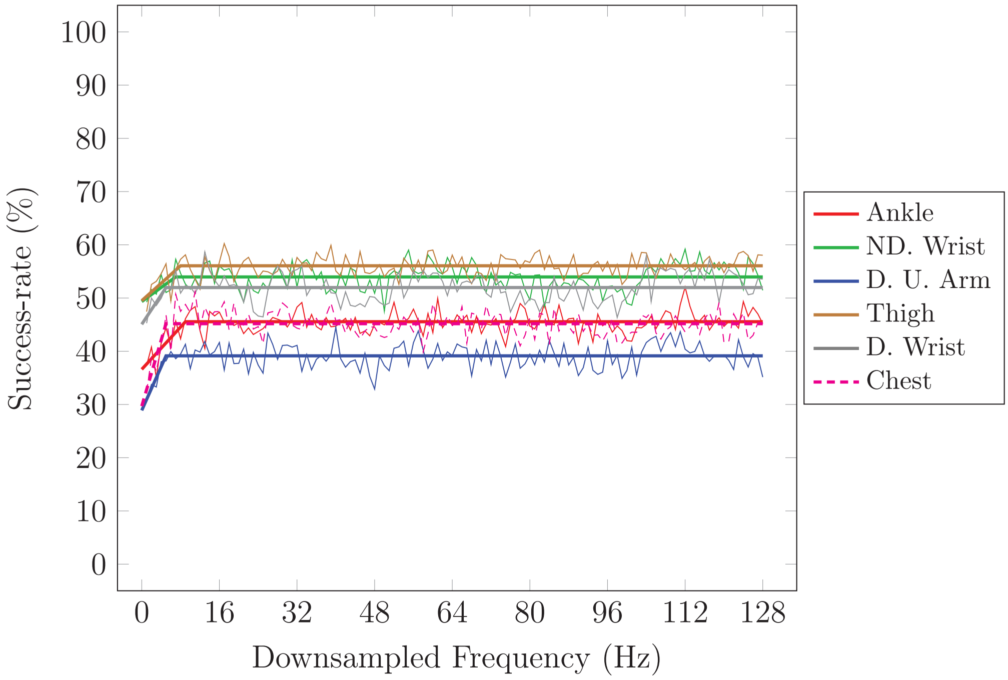
Walking Down Stairs



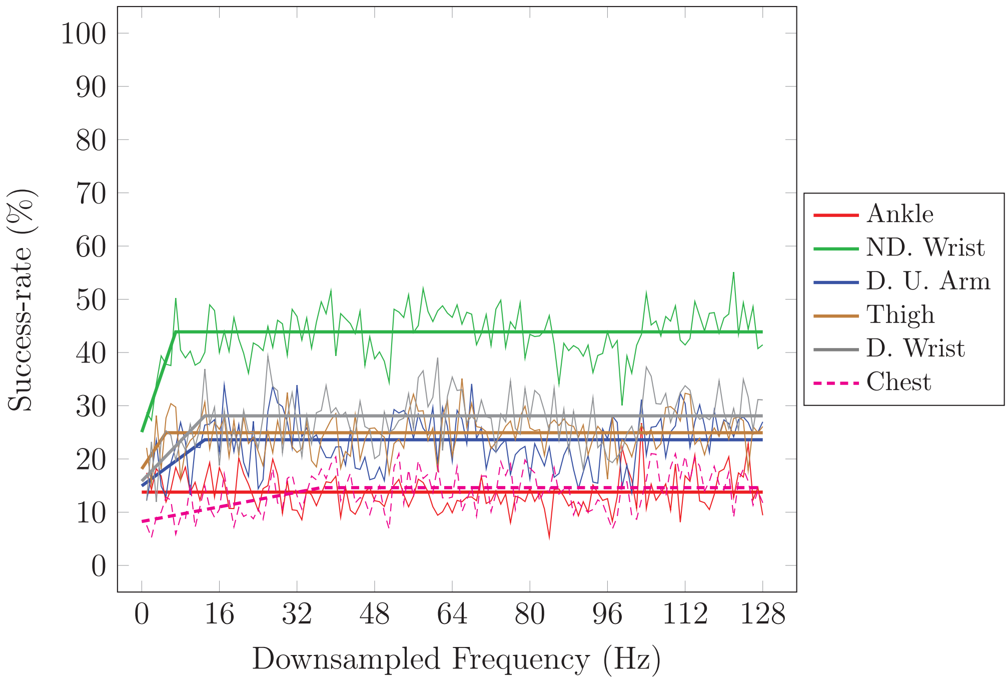
Walking Up Stairs



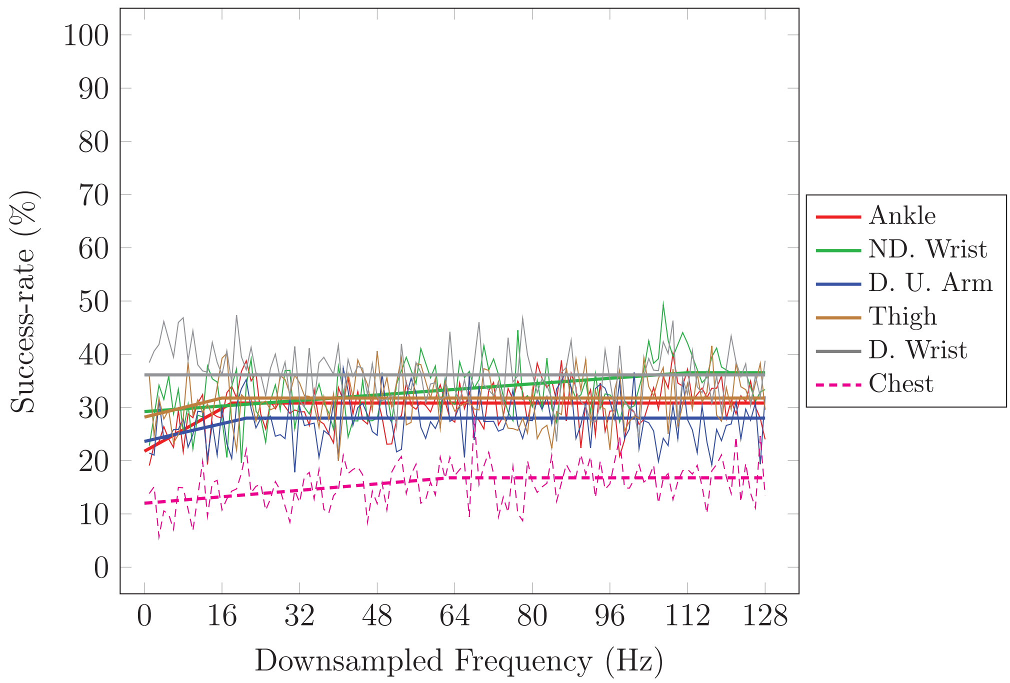
Washing Dishes



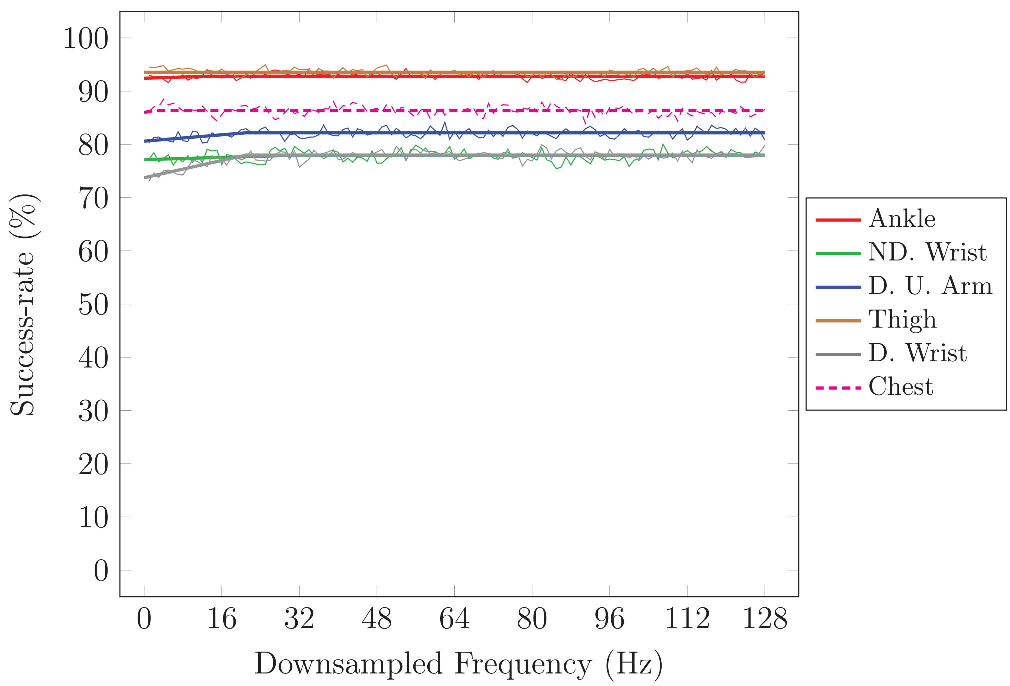
Washing Hands



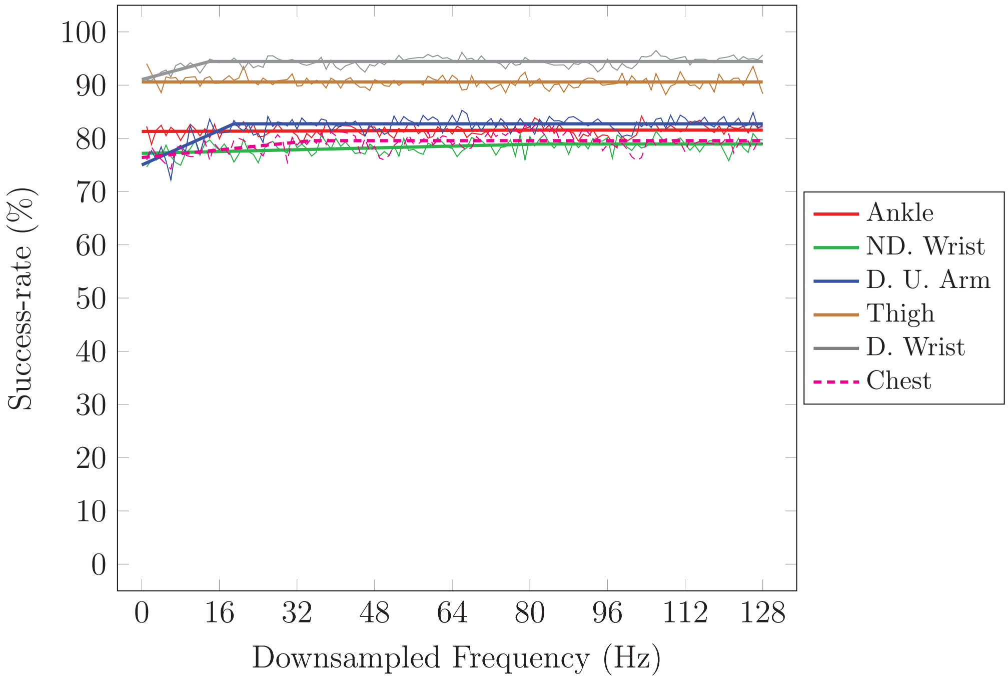
Washing Vegetables



Watching TV

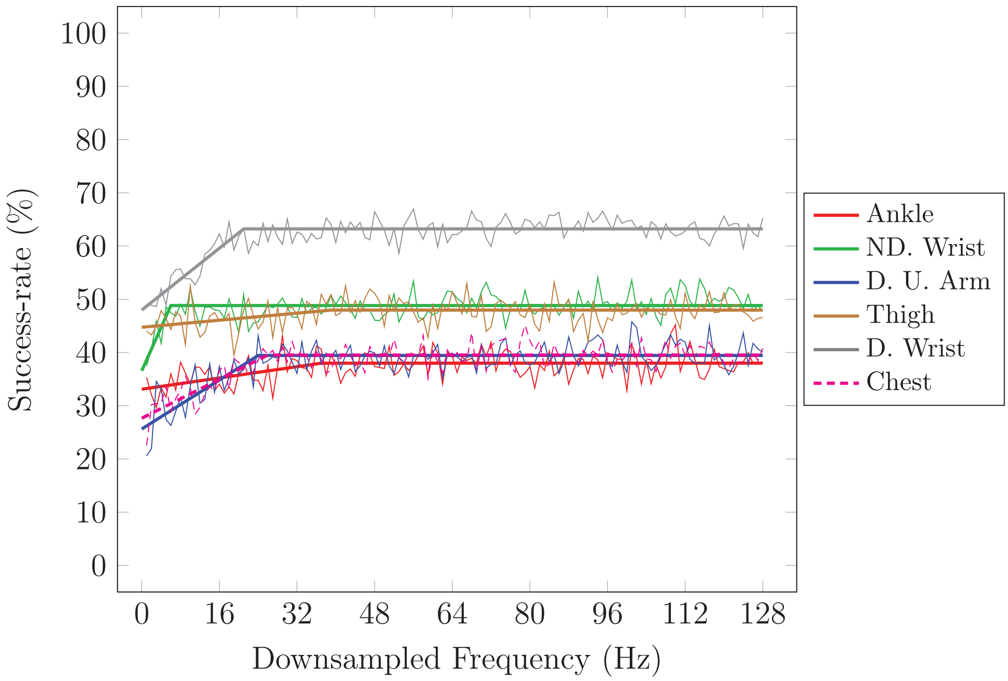


Writing

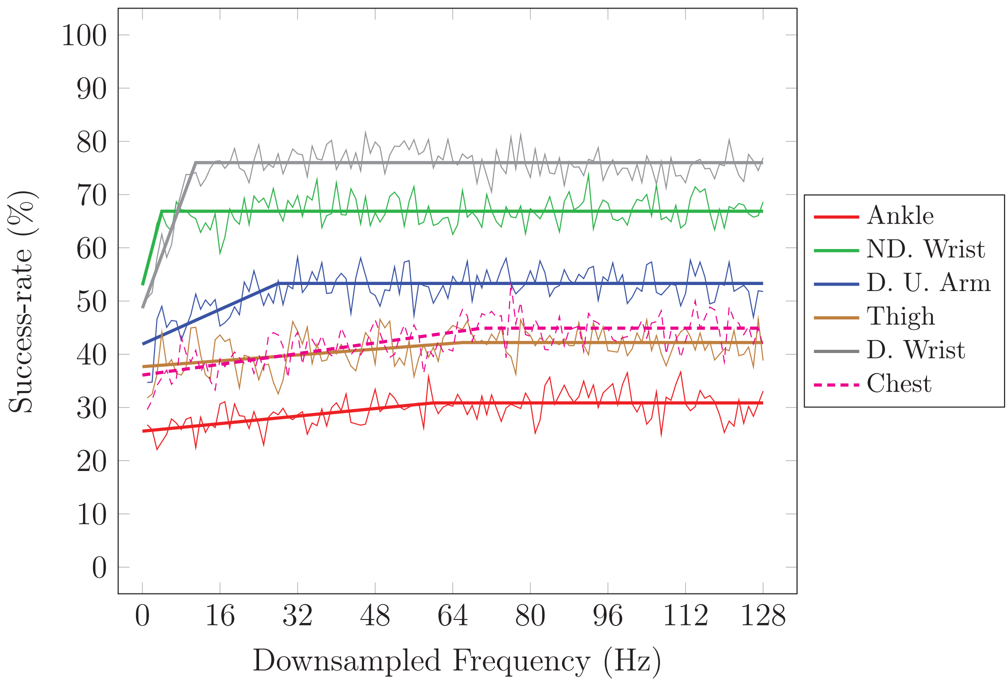


Kwapisz et al. (2011) - Accelerometer

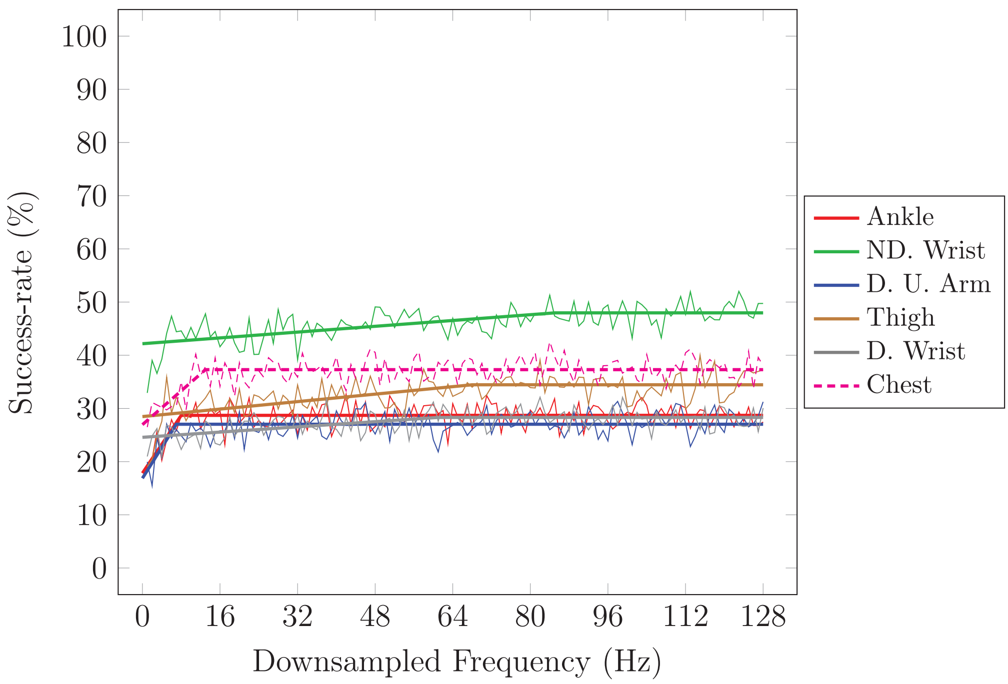
Brushing Teeth



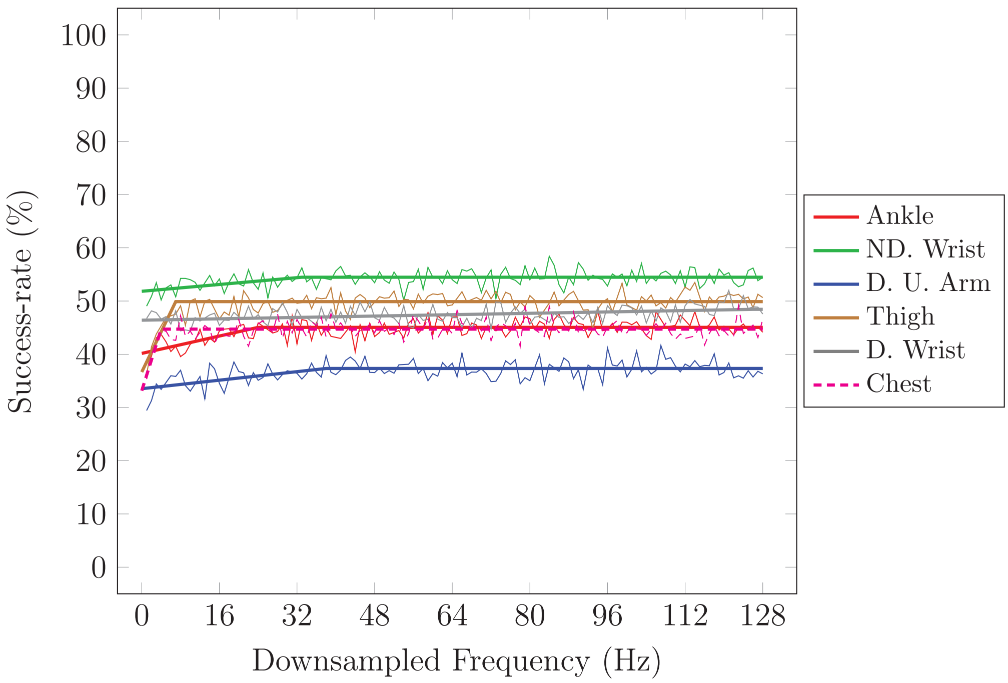
Dicing



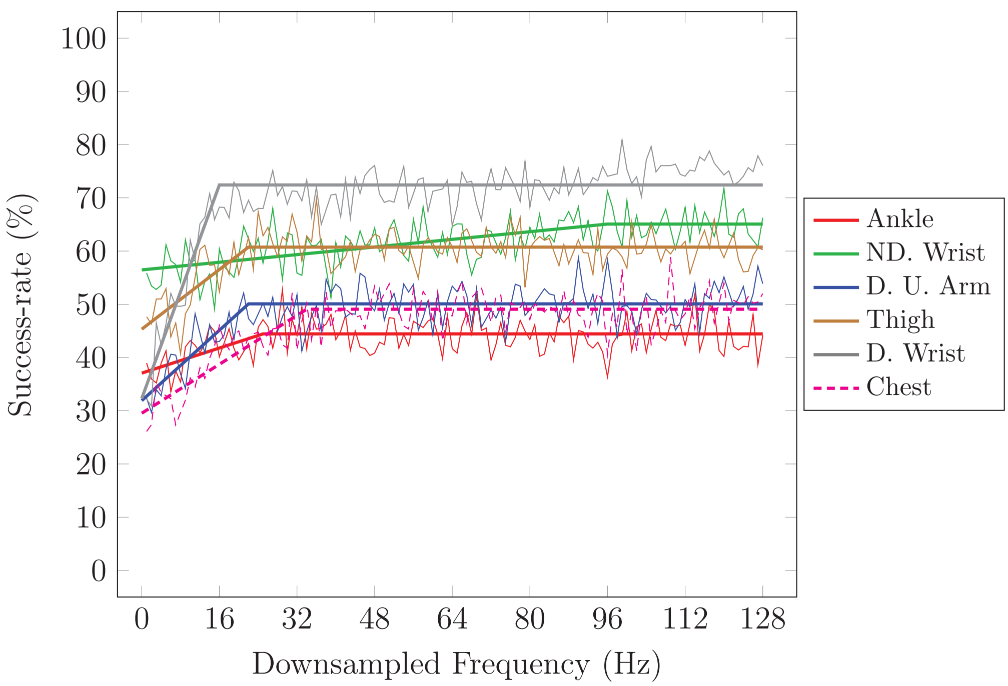
Dusting



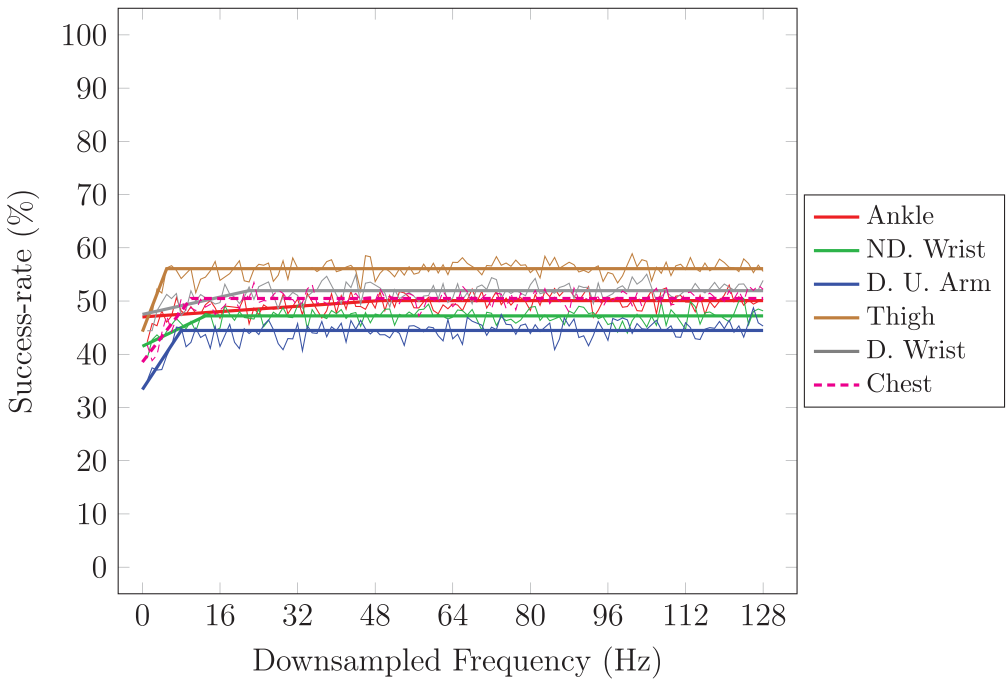
Folding Clothes



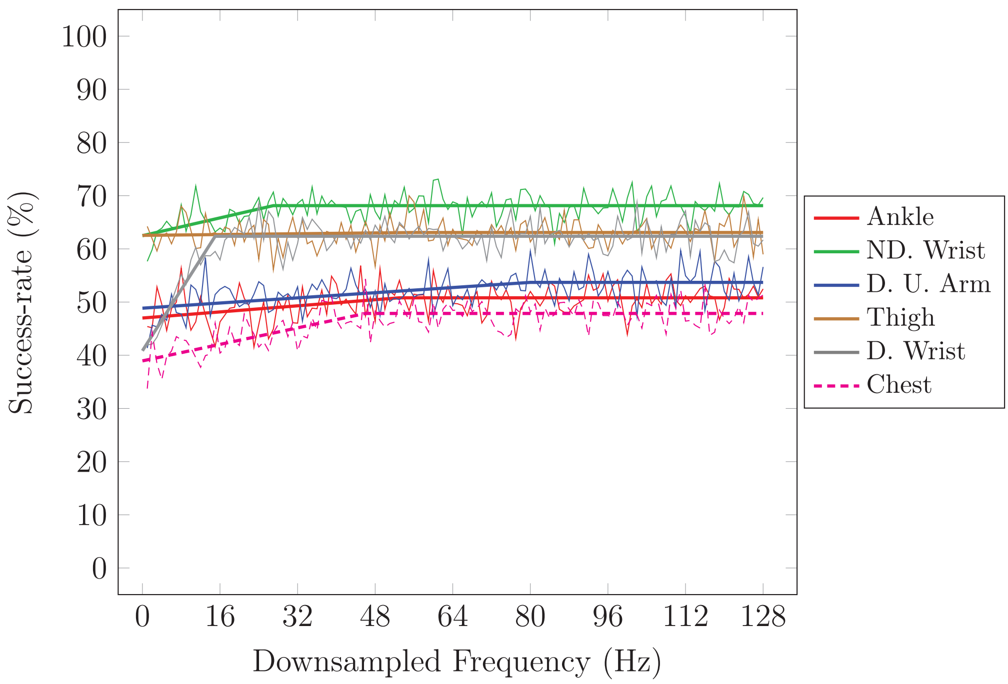
Grating



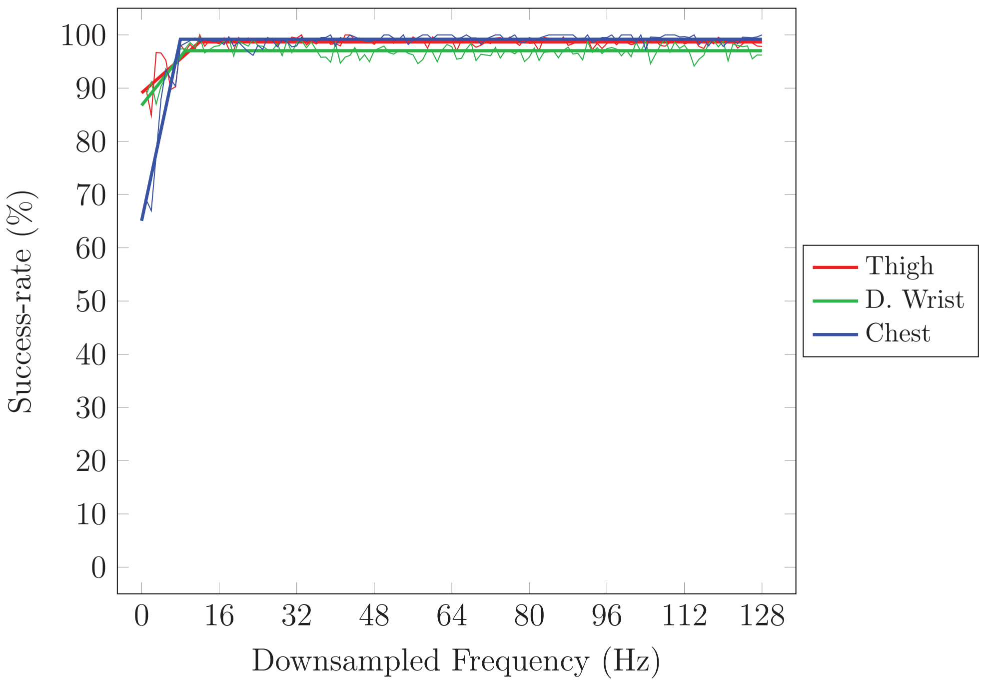
Ironing



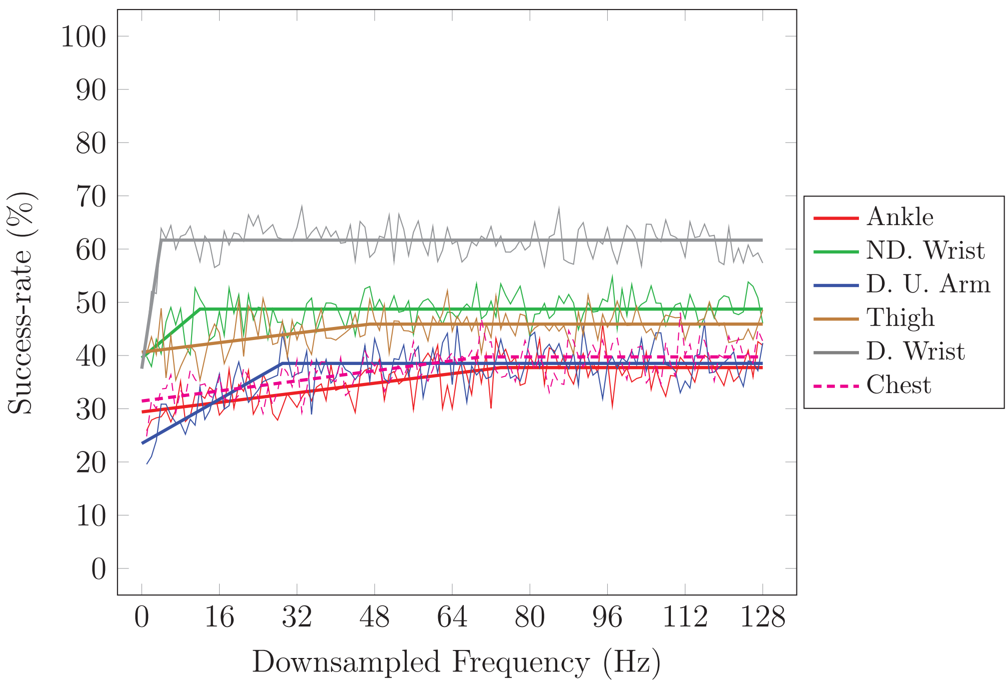
Peeling Vegetables



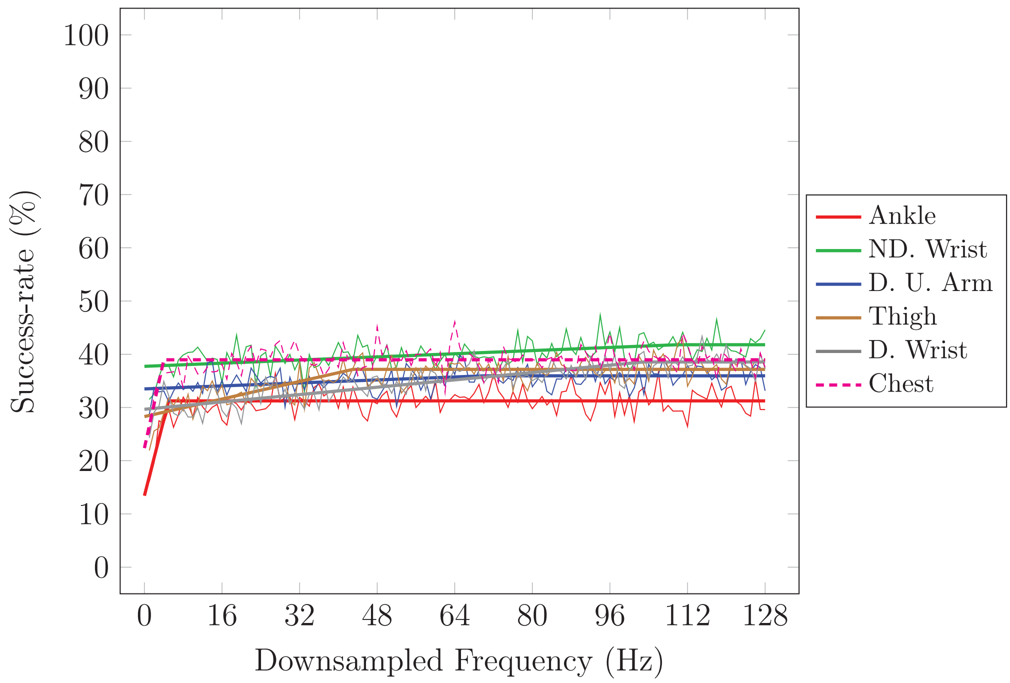
Running



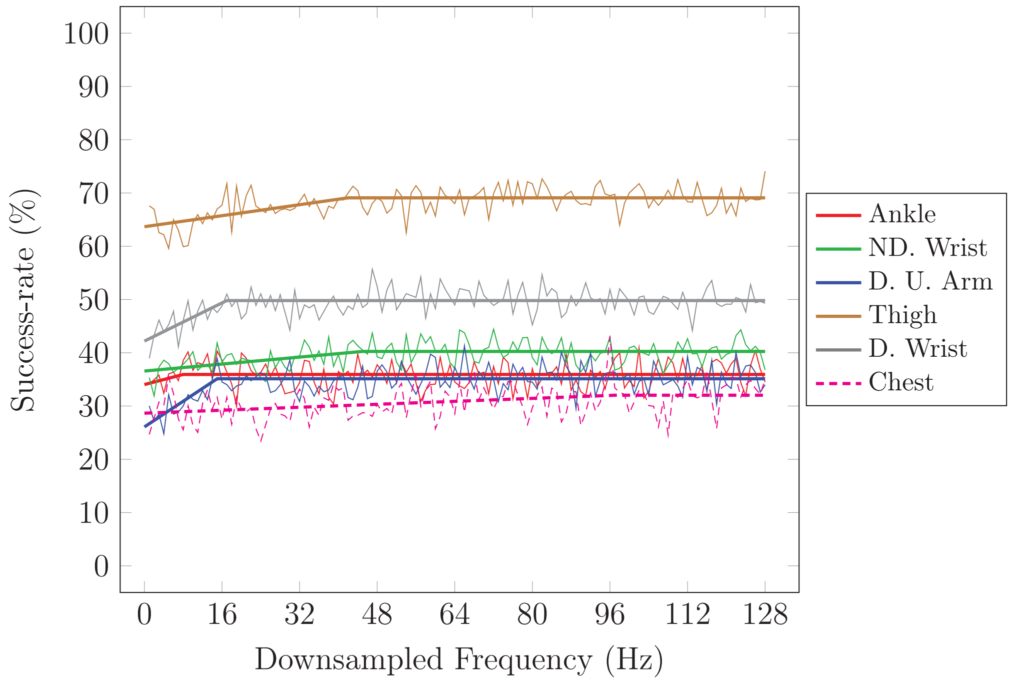
Stiring



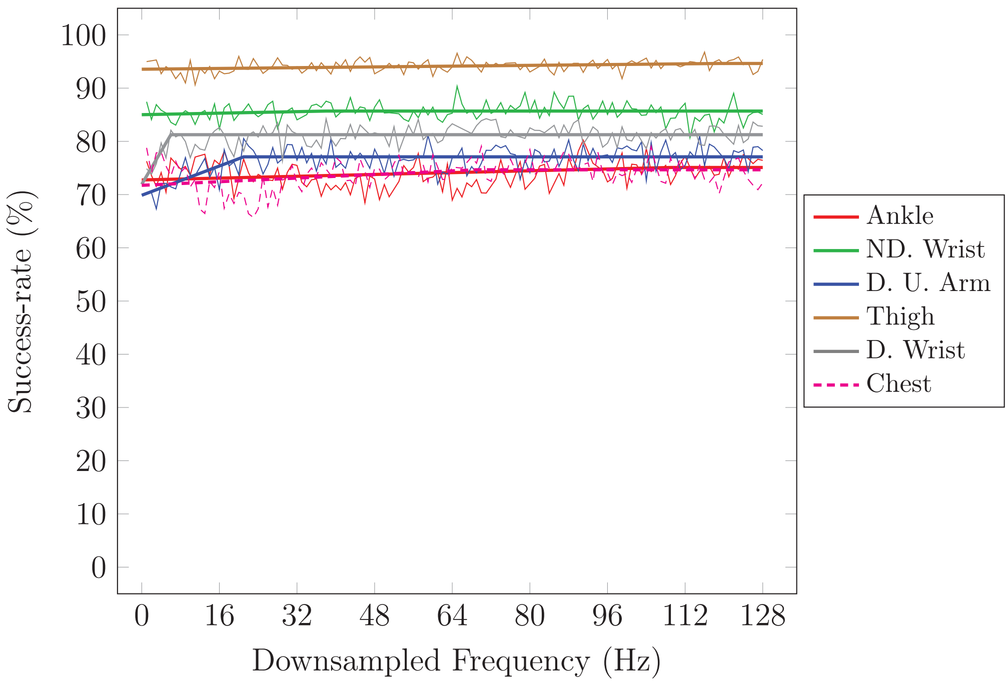
Sweeping



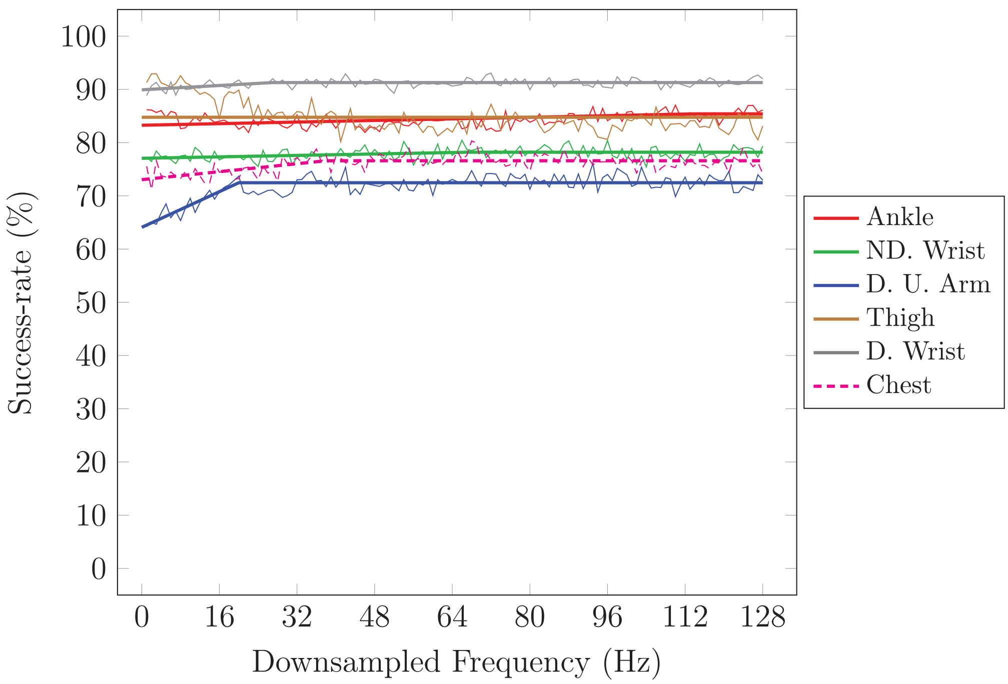
Talking on a Phone



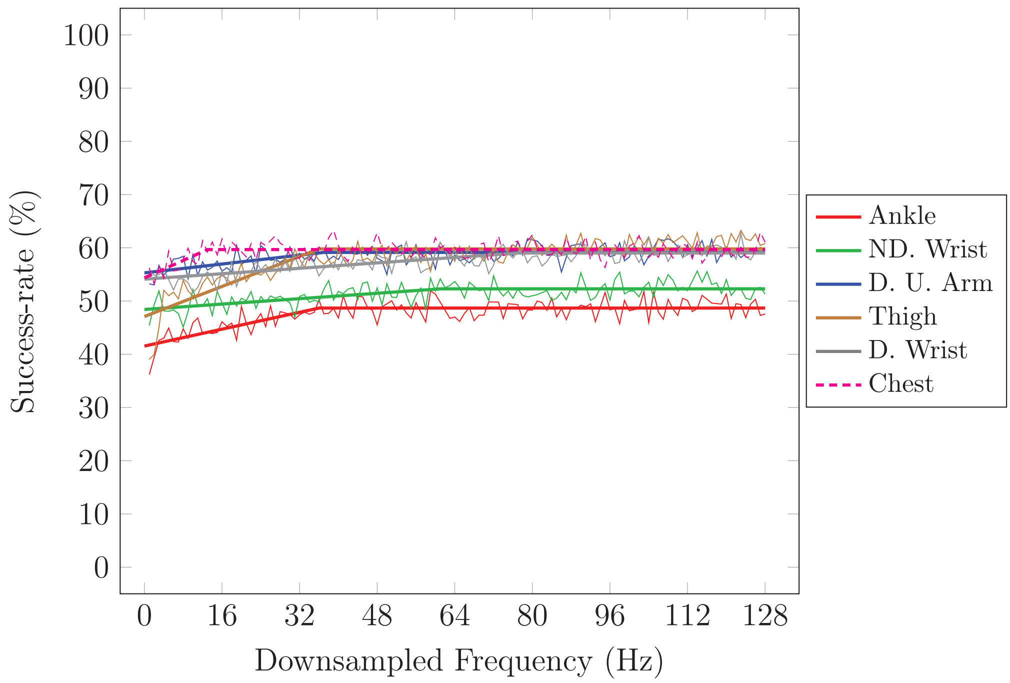
Texting on a Phone



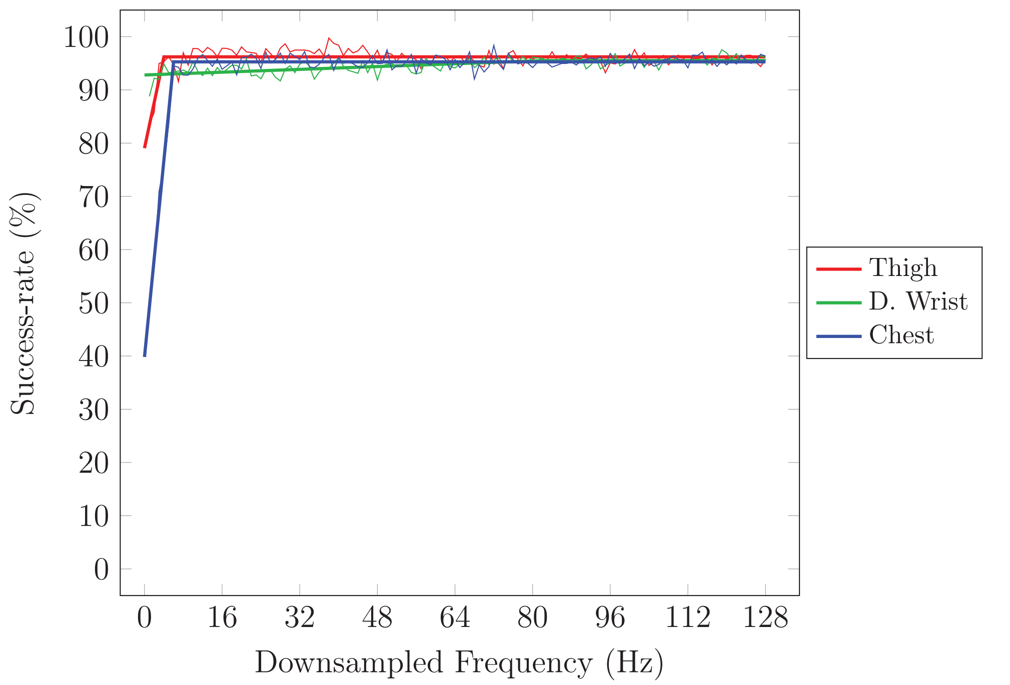
Using a PC



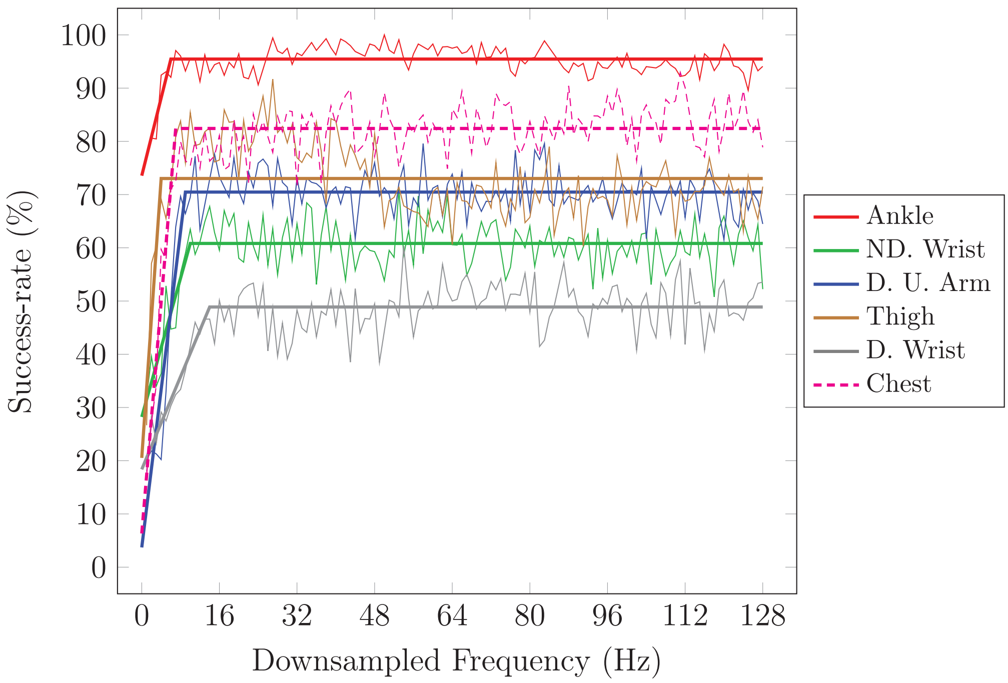
Vacuuming



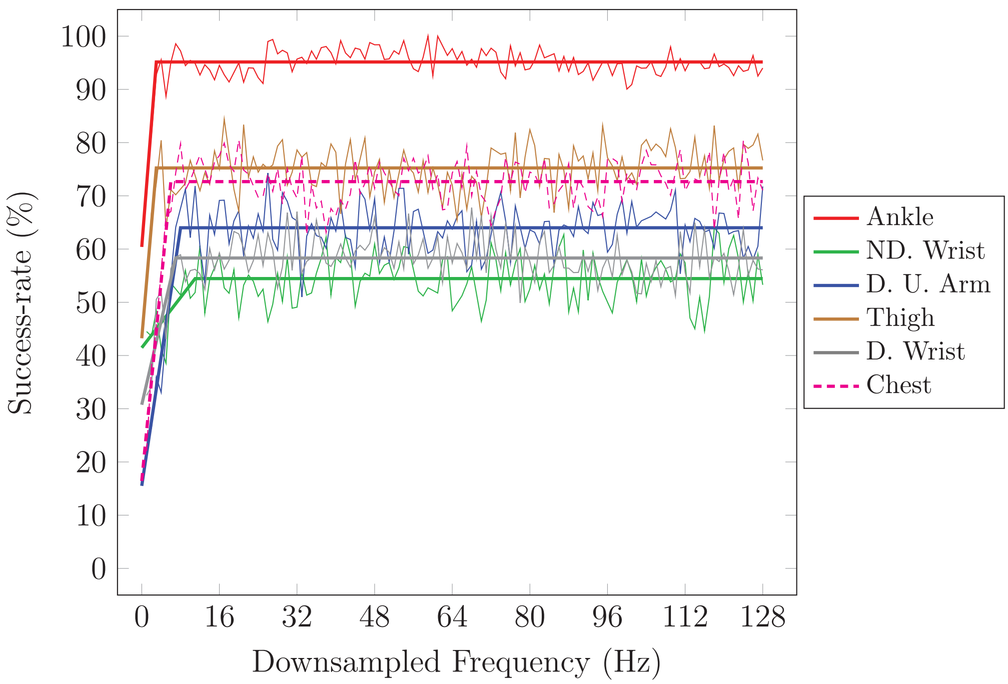
Walking (Flat Ground)



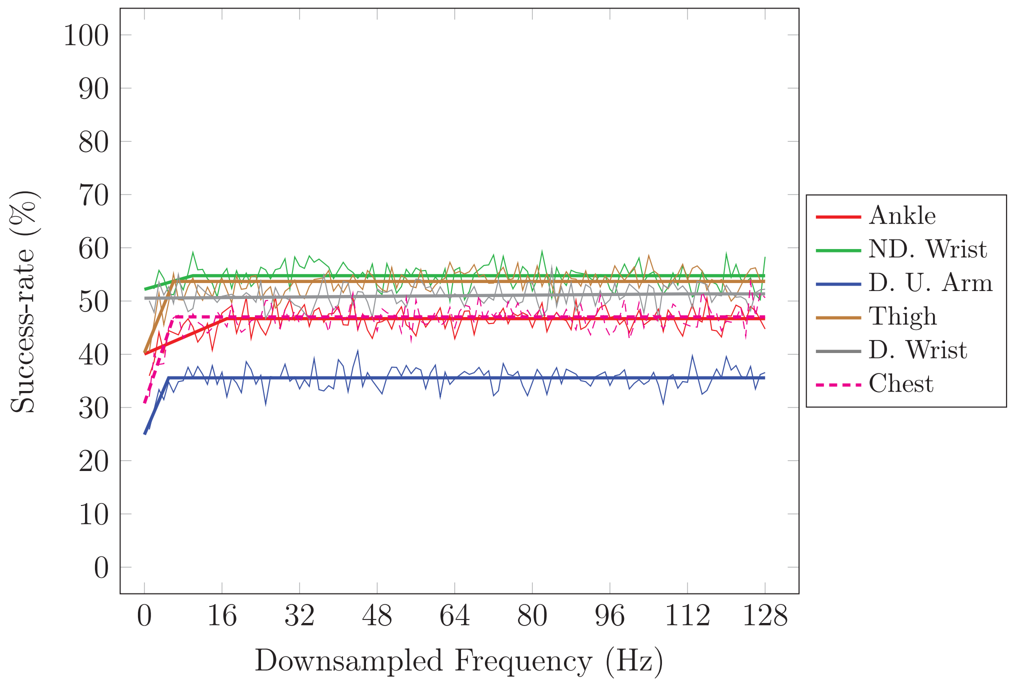
Walking Down Stairs



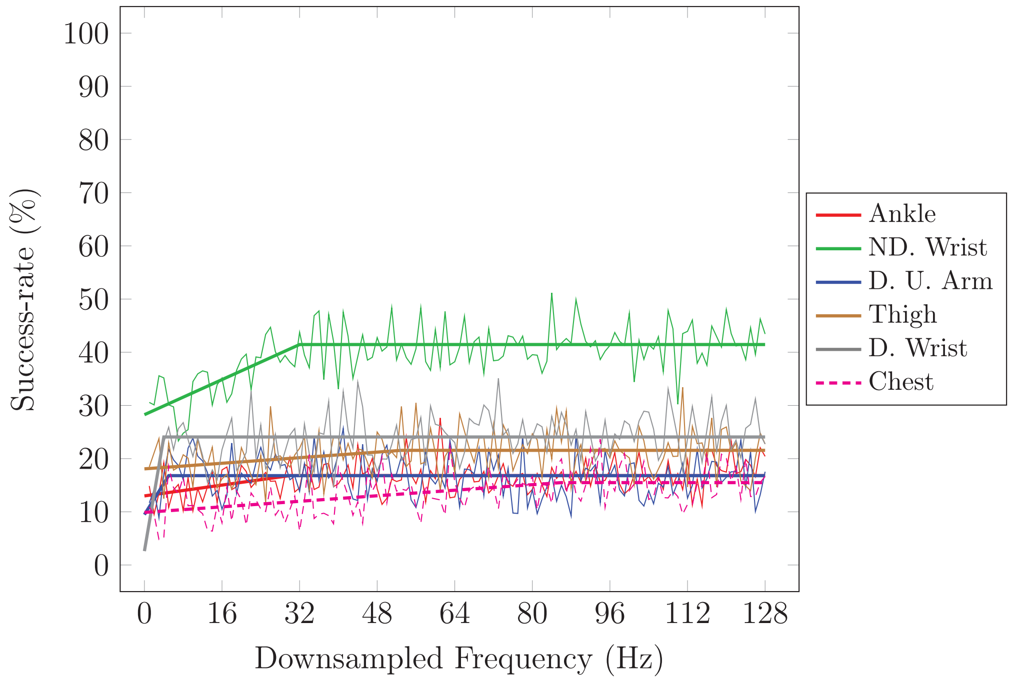
Walking Up Stairs



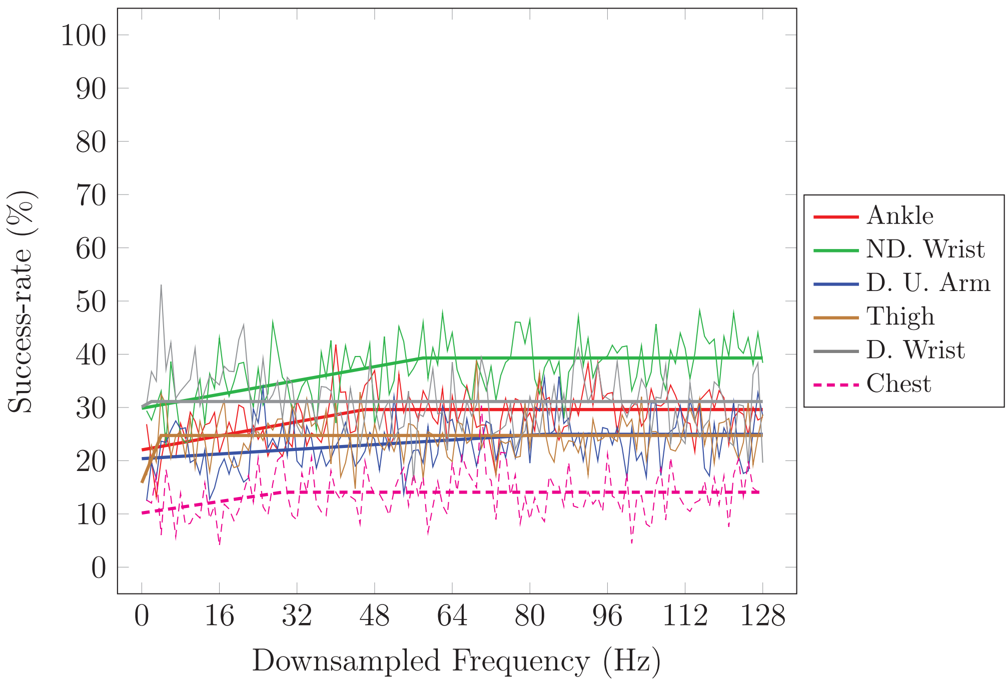
Washing Dishes



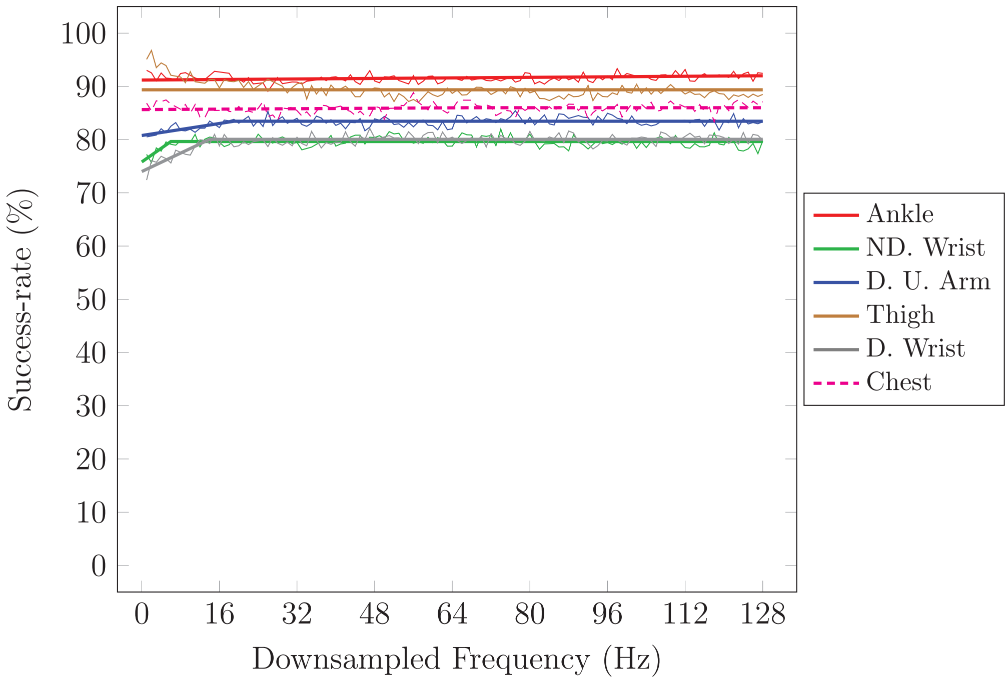
Washing Hands



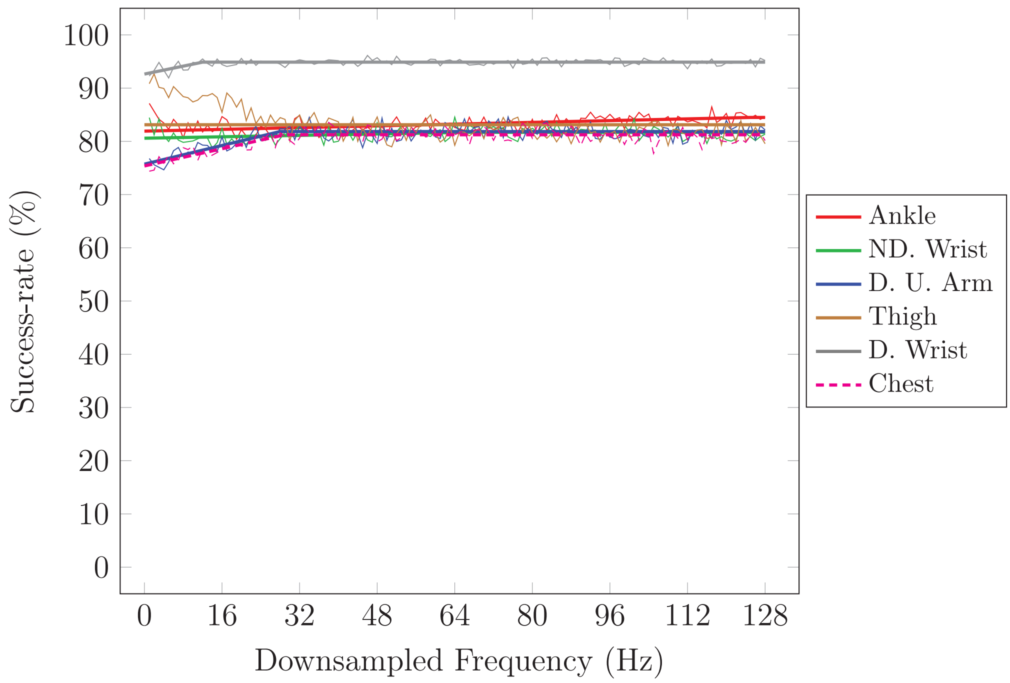
Washing Vegetables



Watching TV

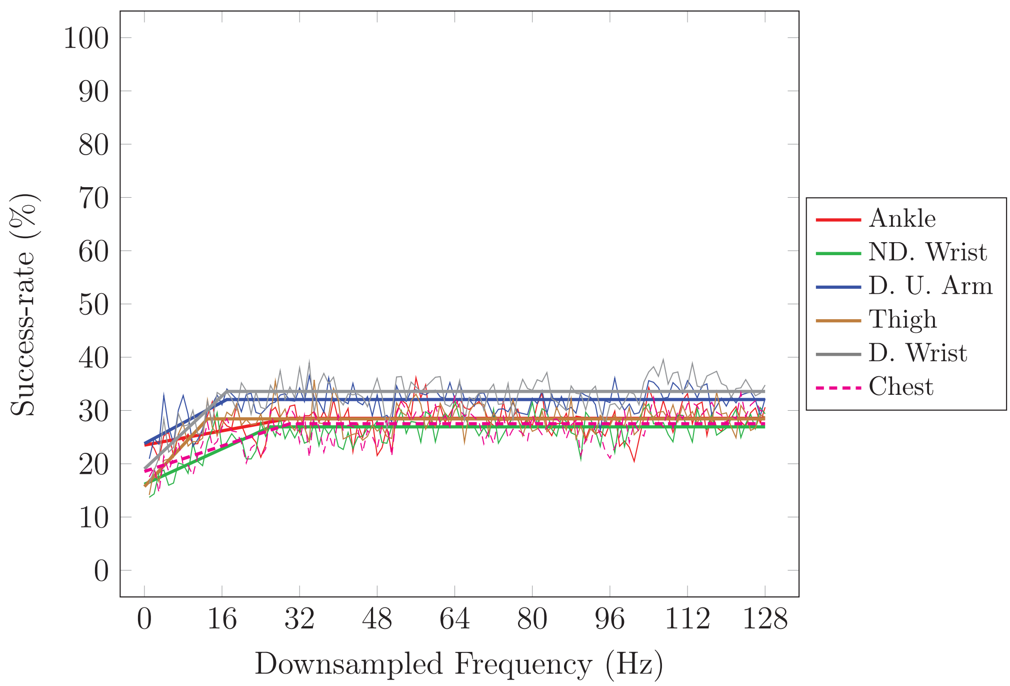


Writing

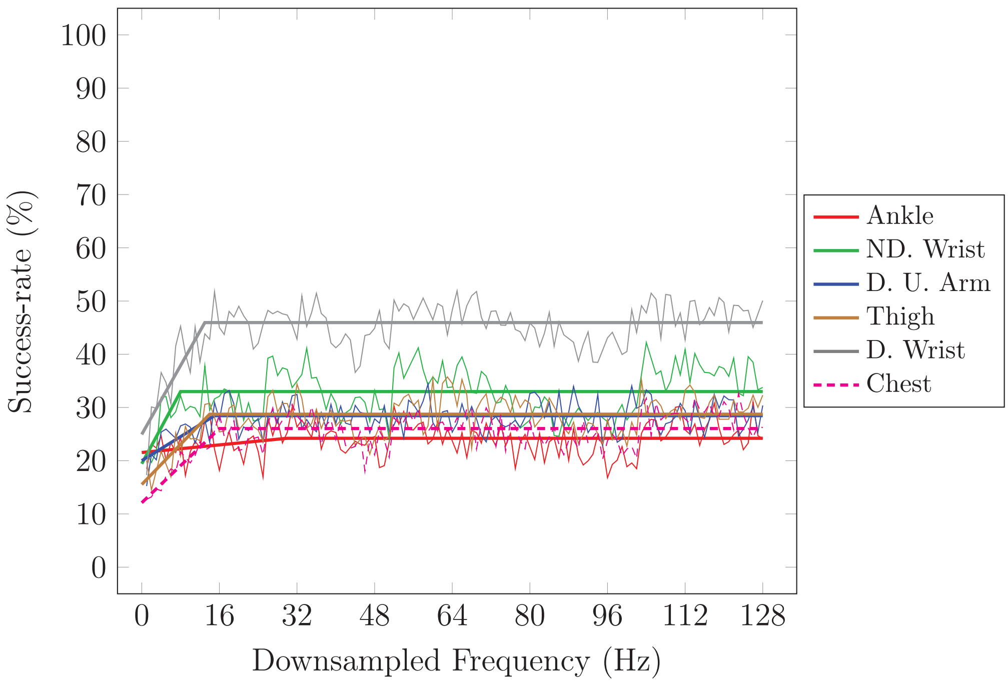


Bao and Intille (2004) - Gyroscope

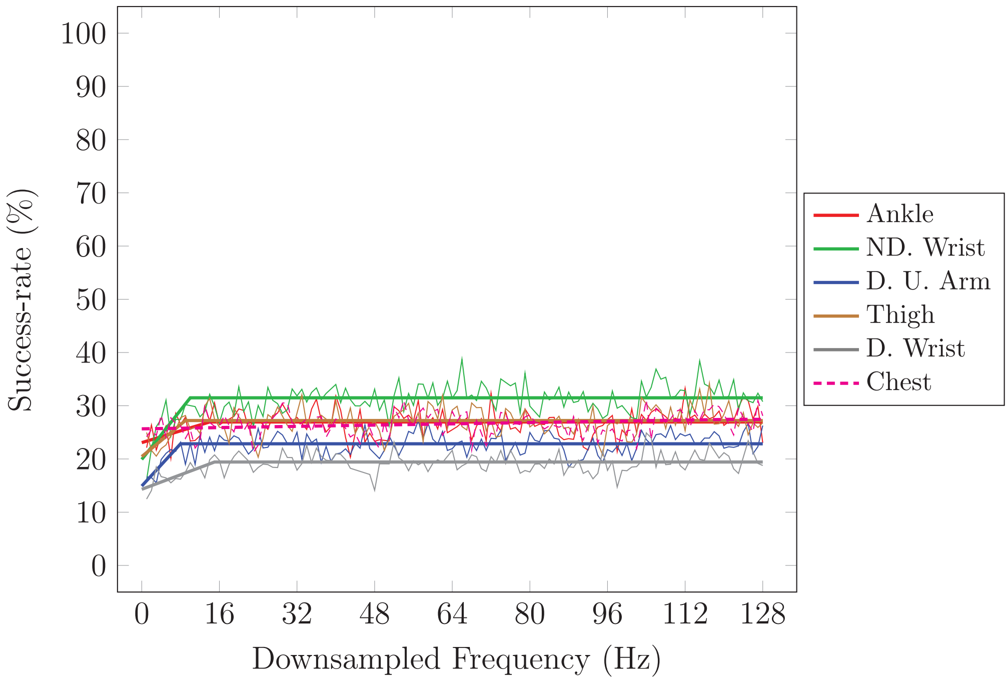
Brushing Teeth



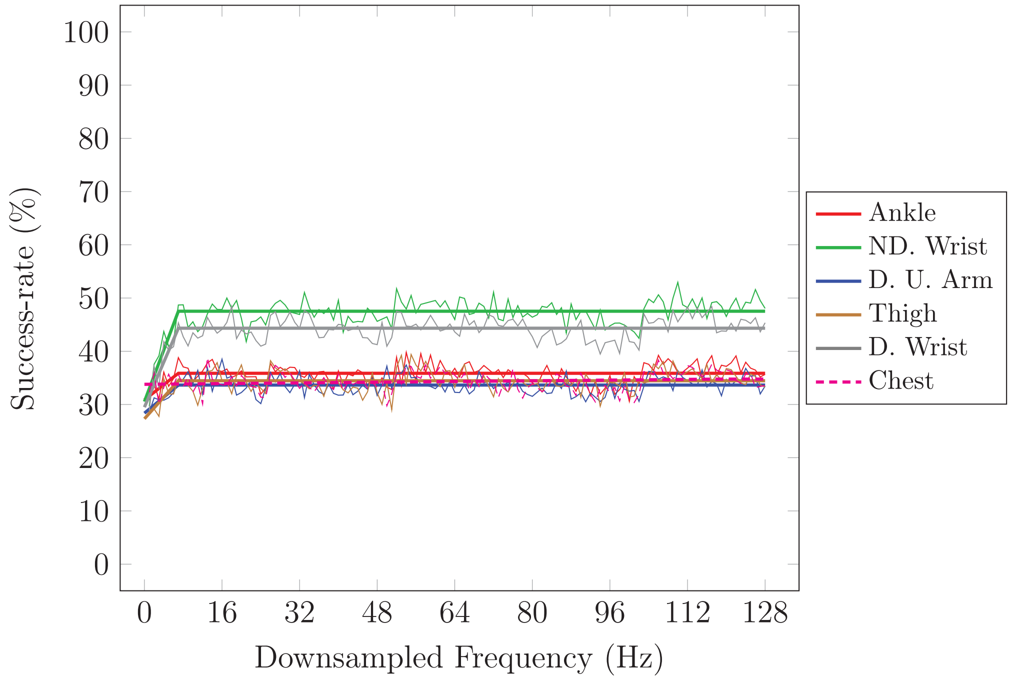
Dicing



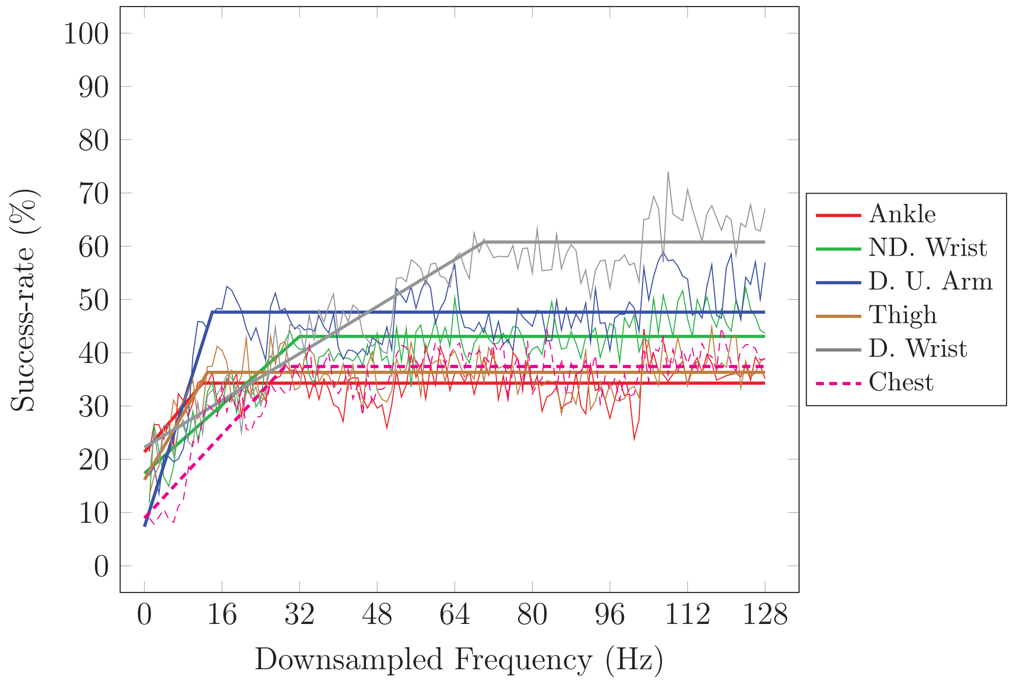
Dusting



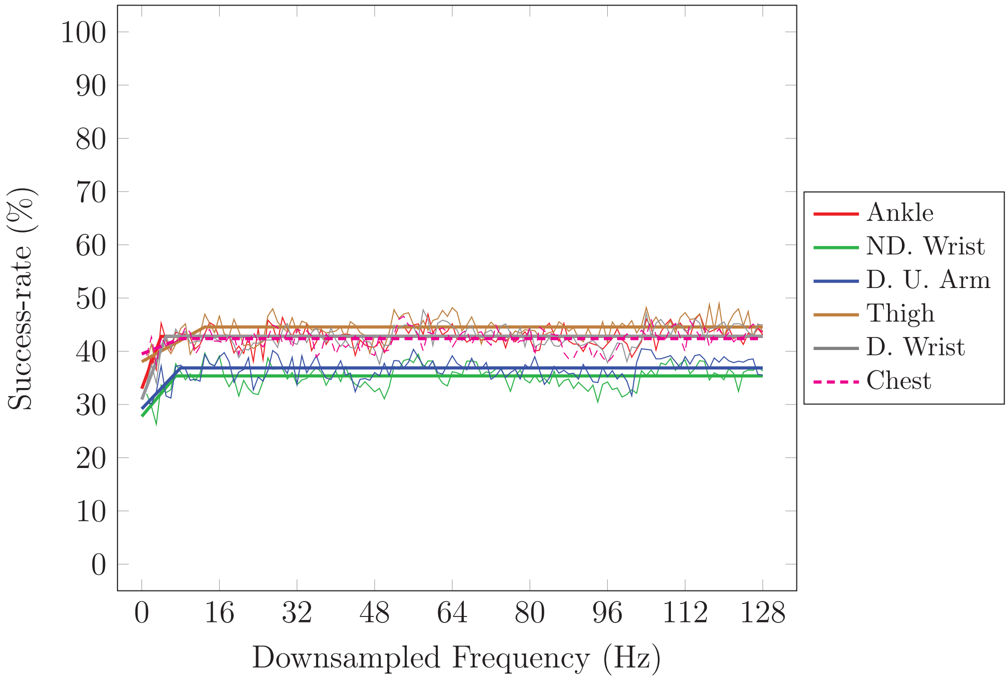
Folding Clothes



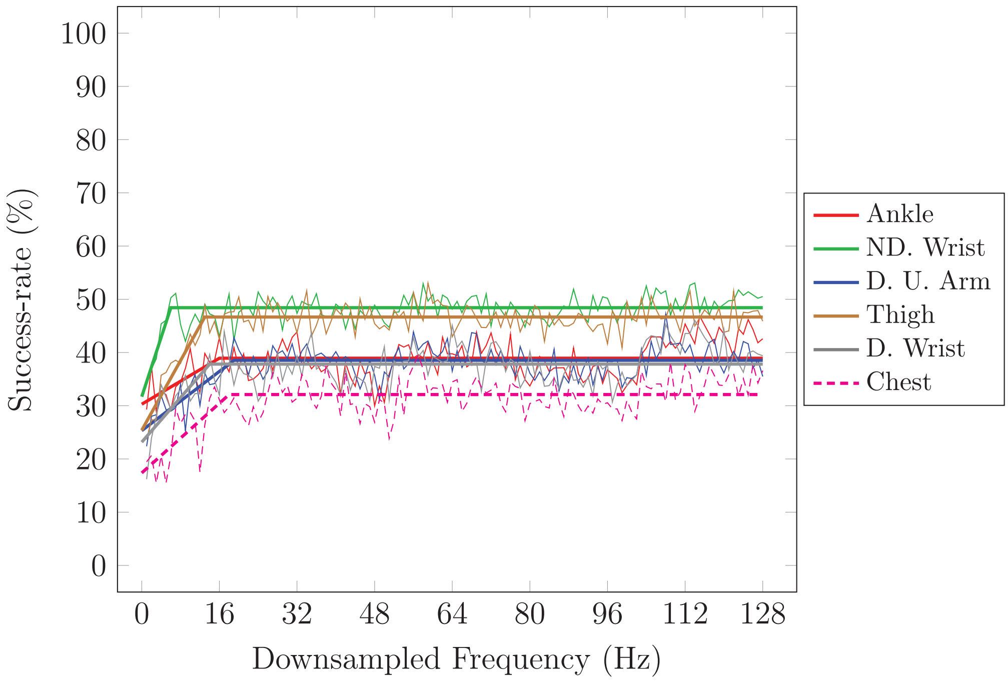
Grating



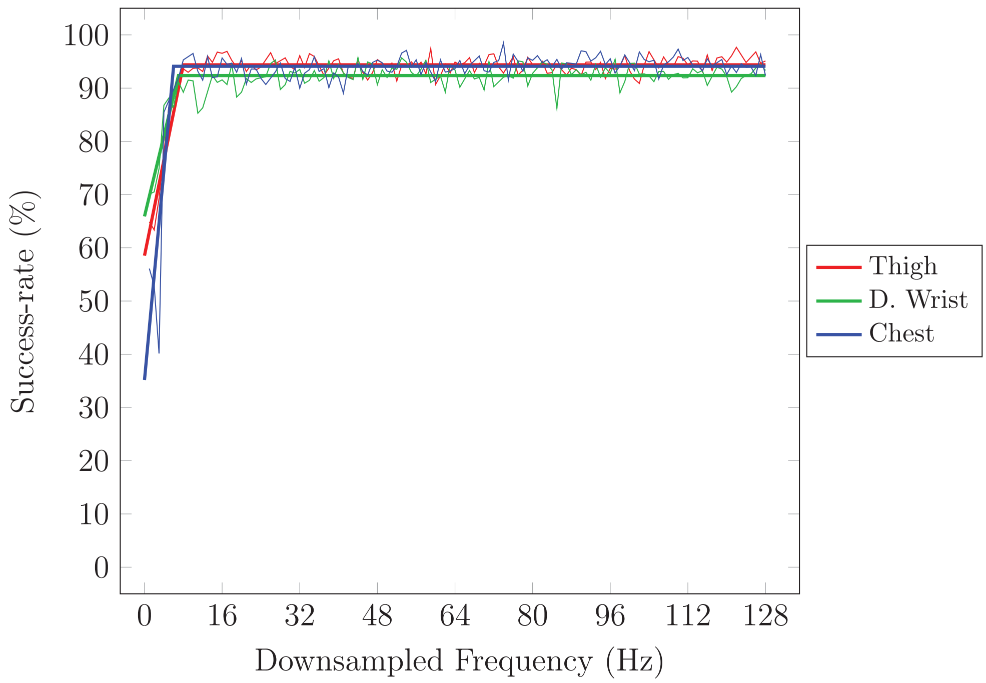
Ironing



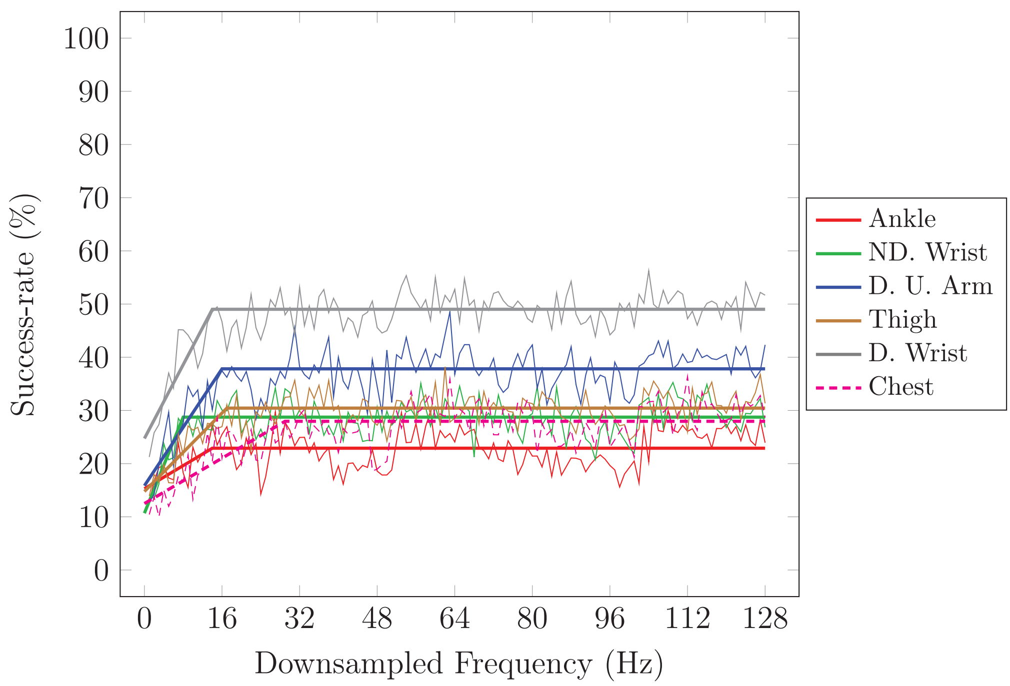
Peeling Vegetables



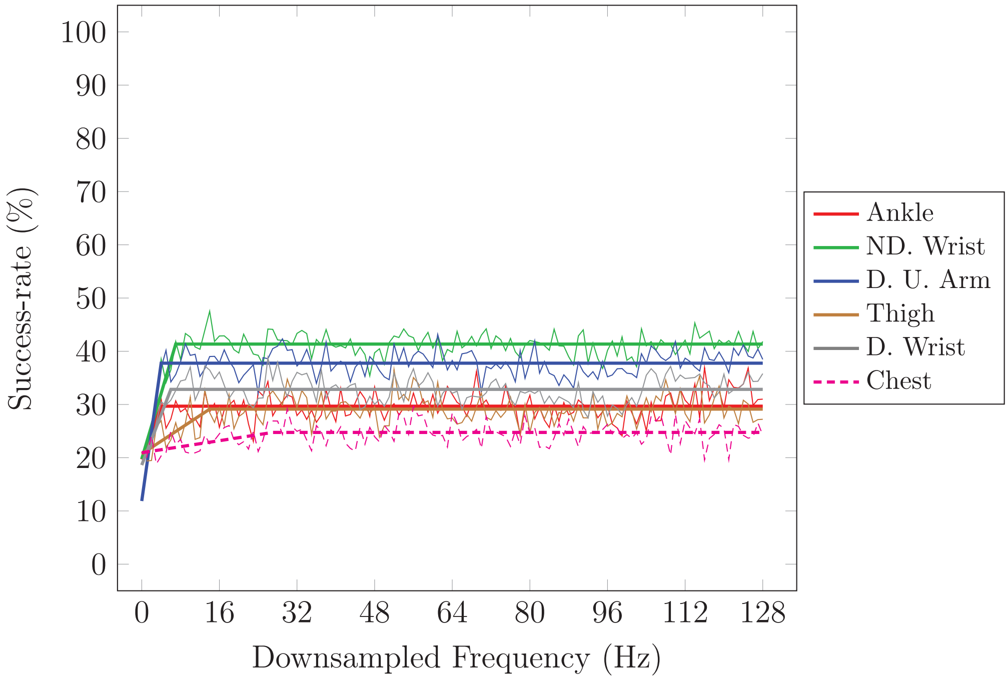
Running



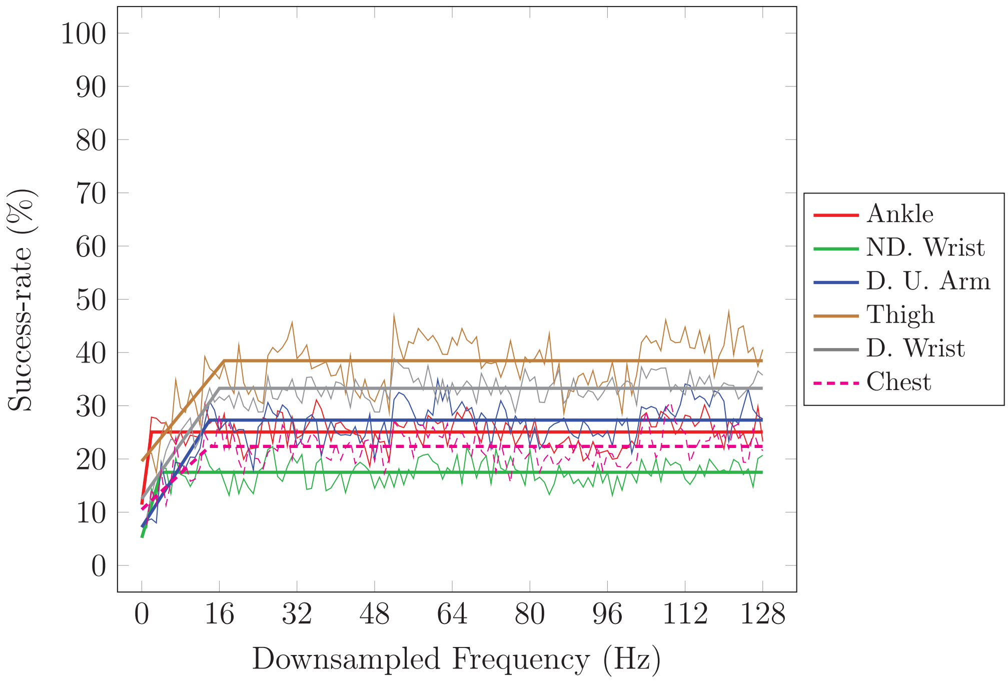
Stiring



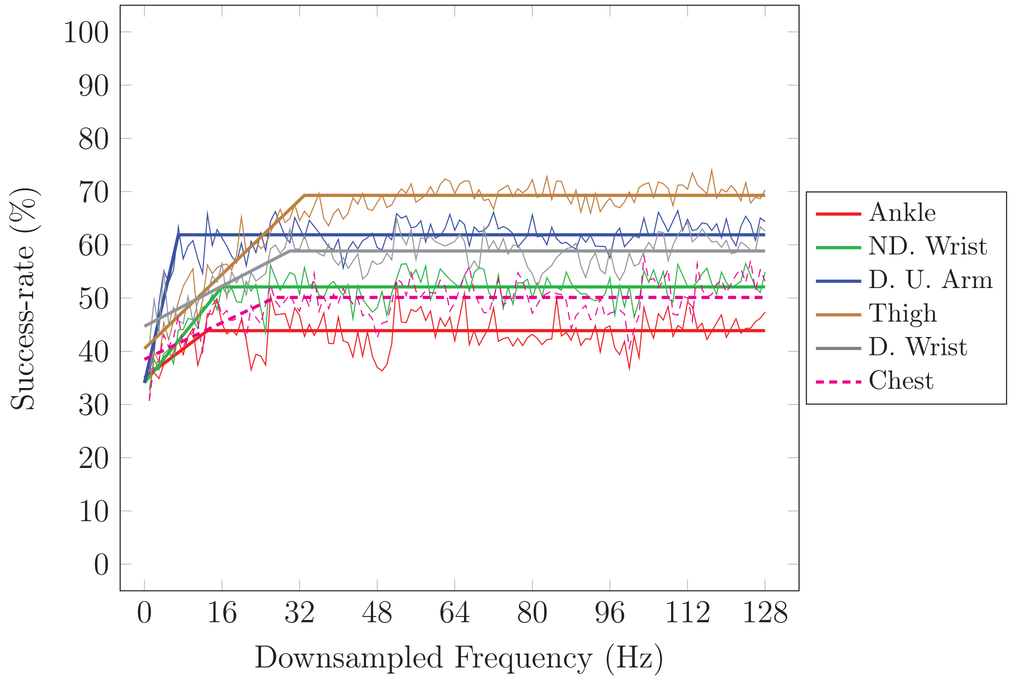
Sweeping



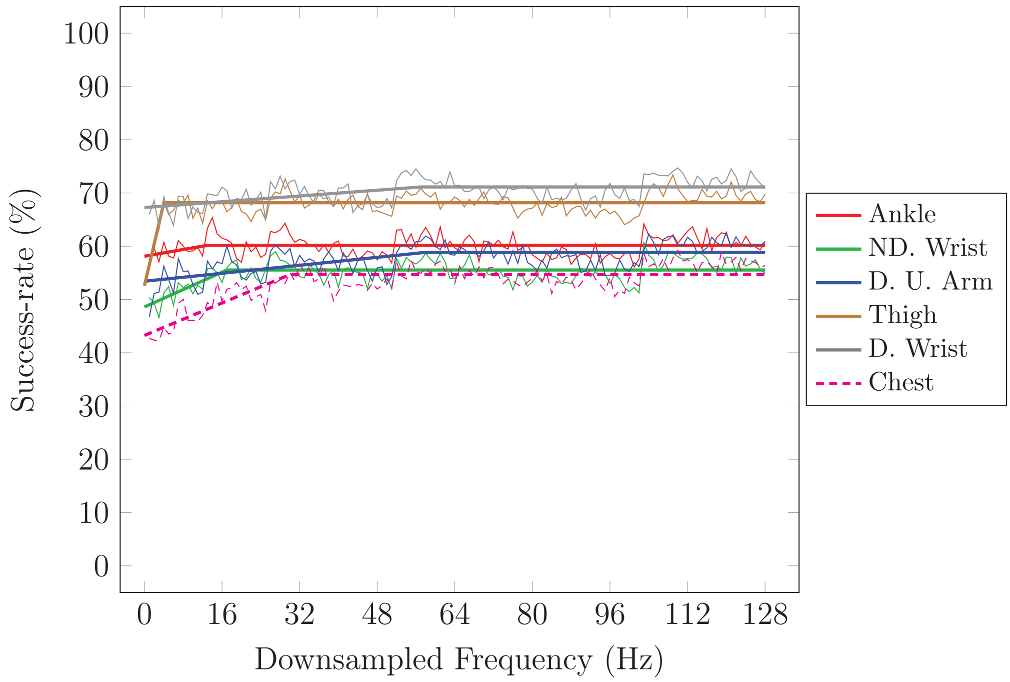
Talking on a Phone



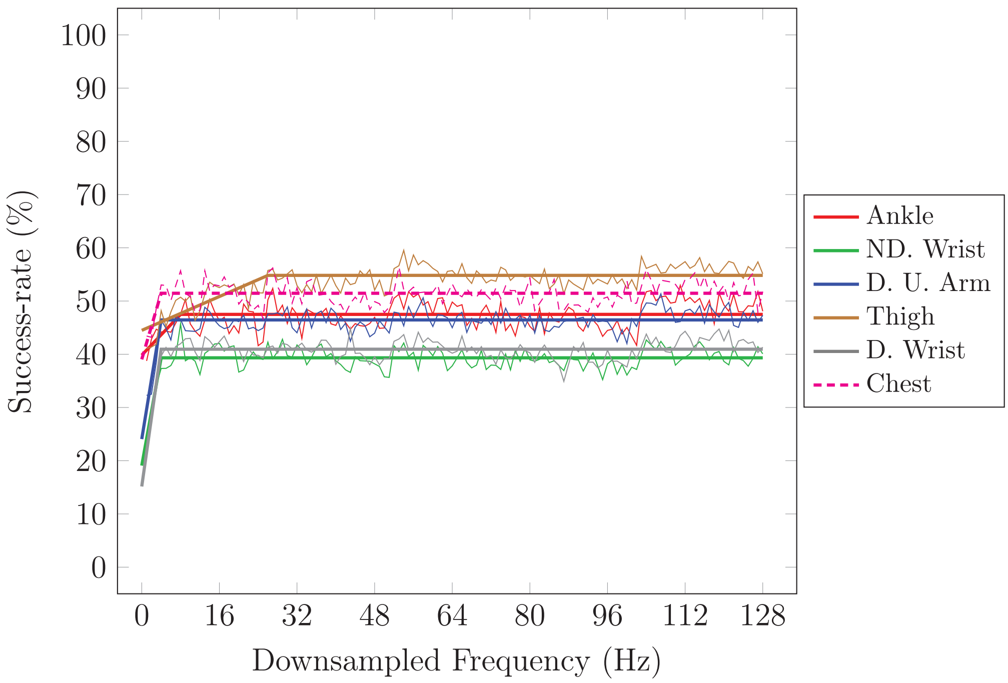
Texting on a Phone



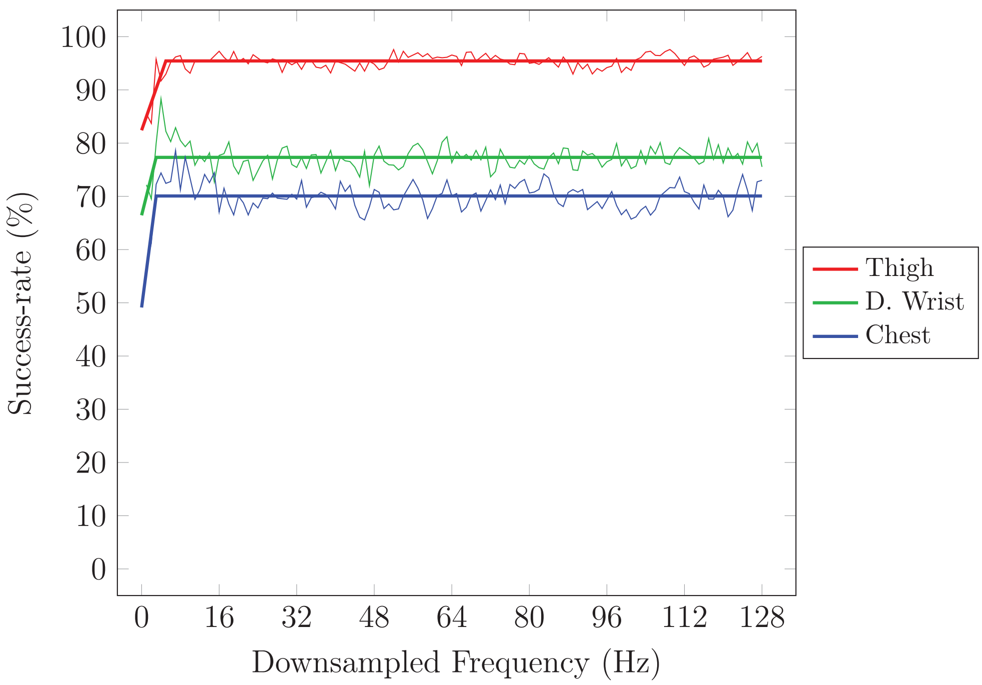
Using a PC



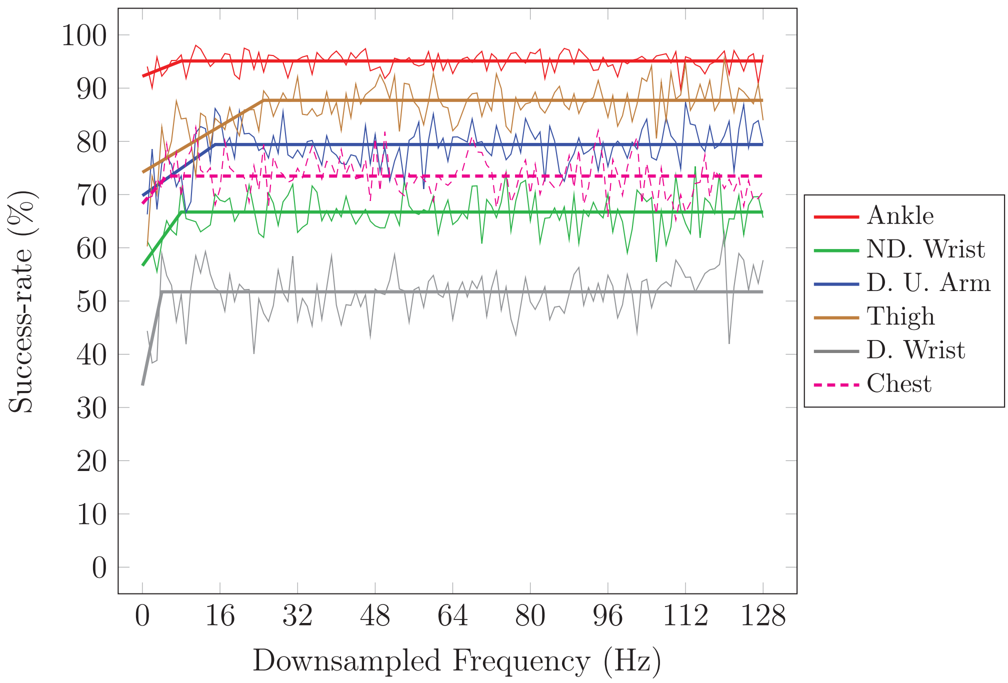
Vacuuming



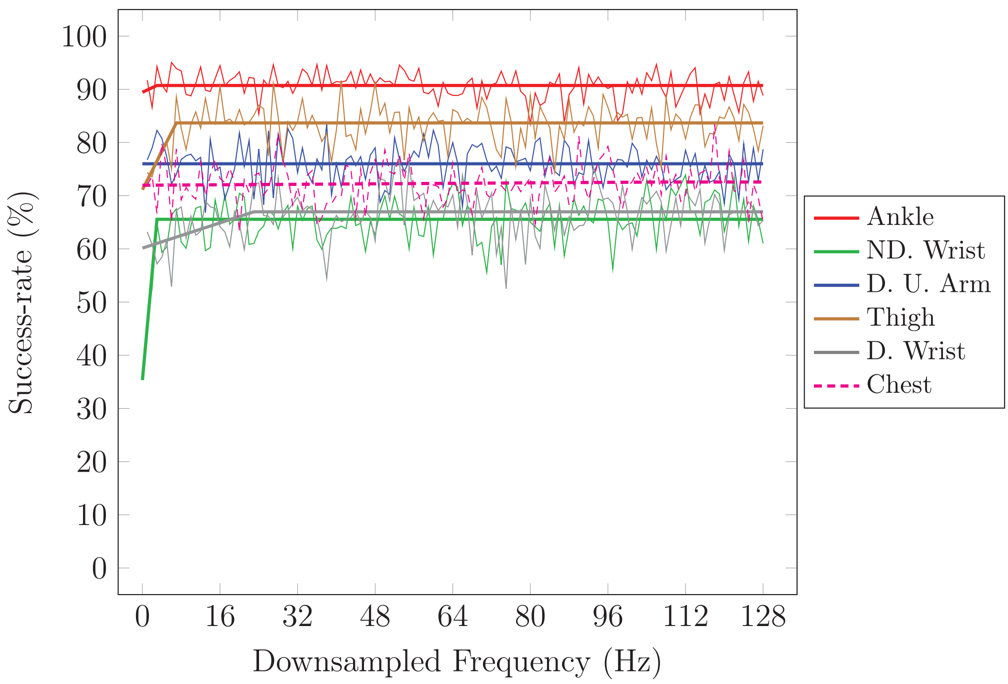
Walking (Flat Ground)



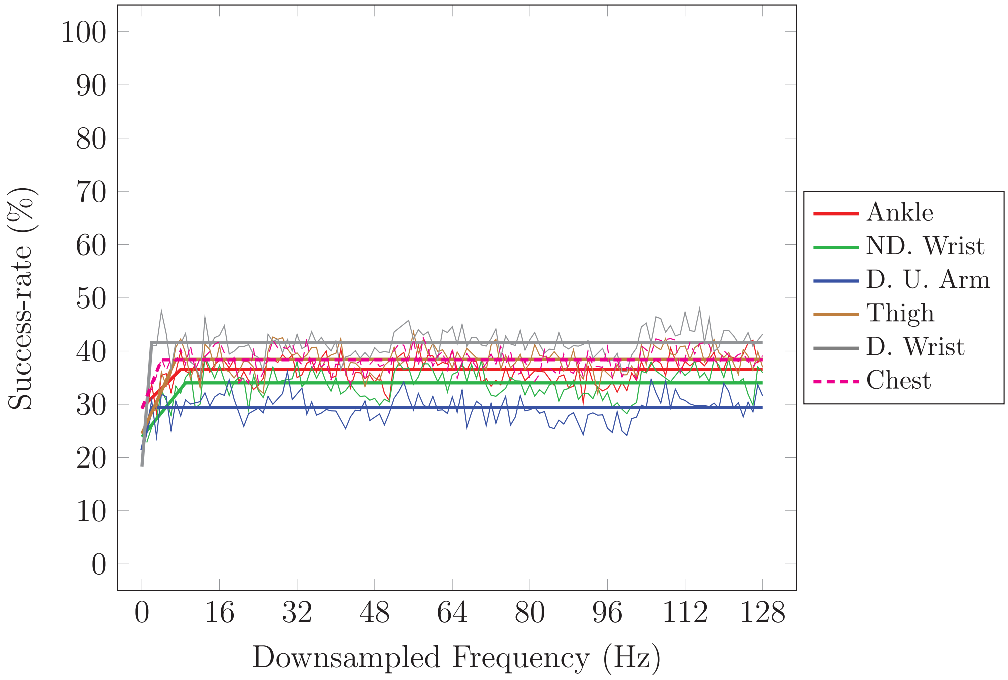
Walking Down Stairs



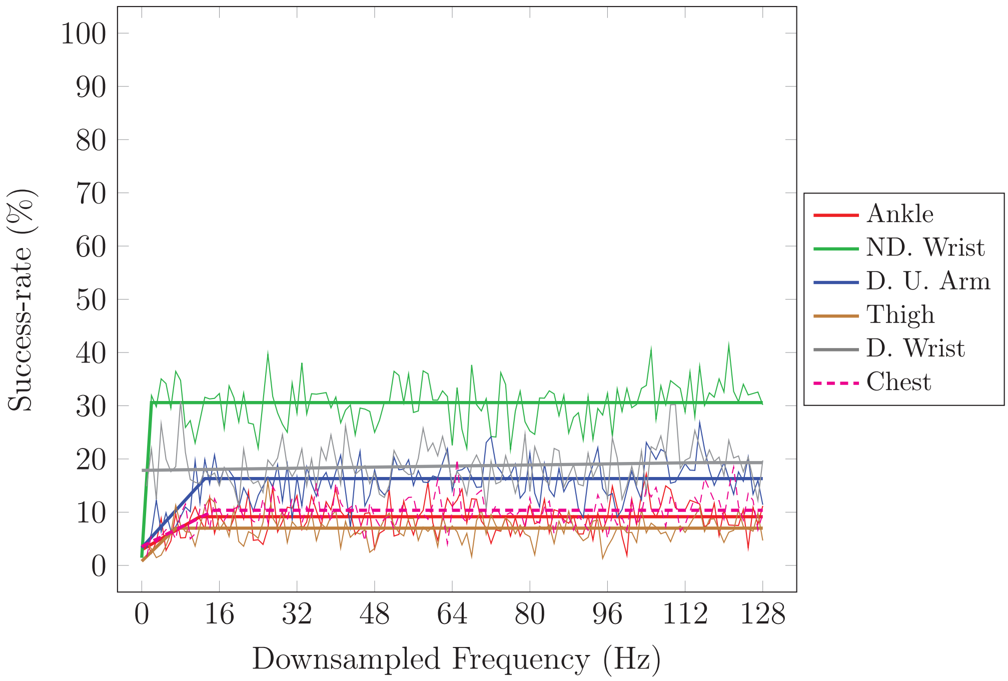
Walking Up Stairs



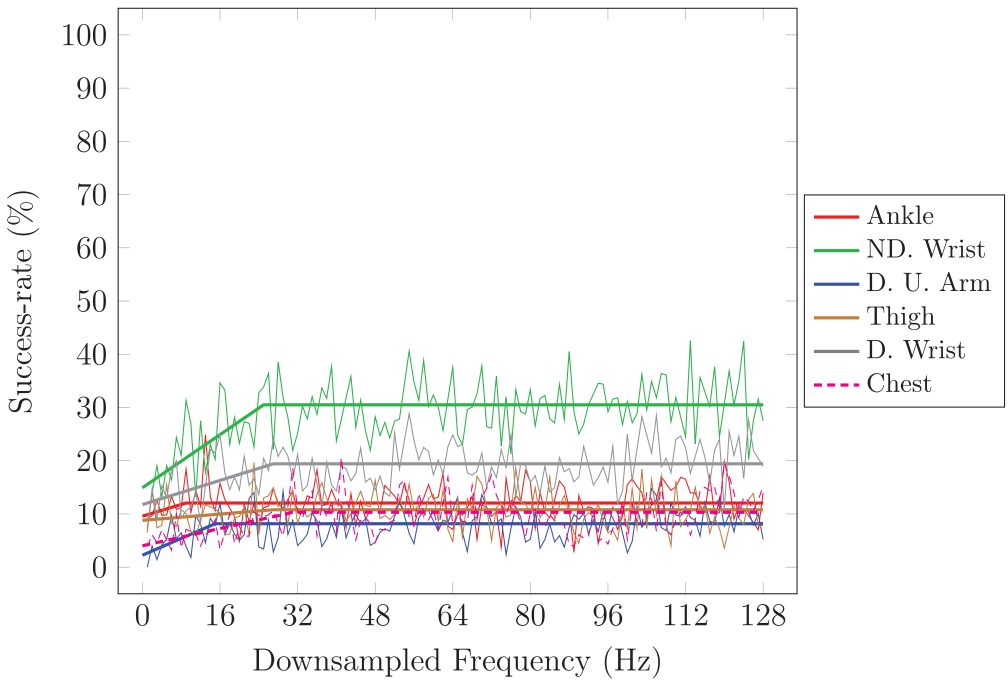
Washing Dishes



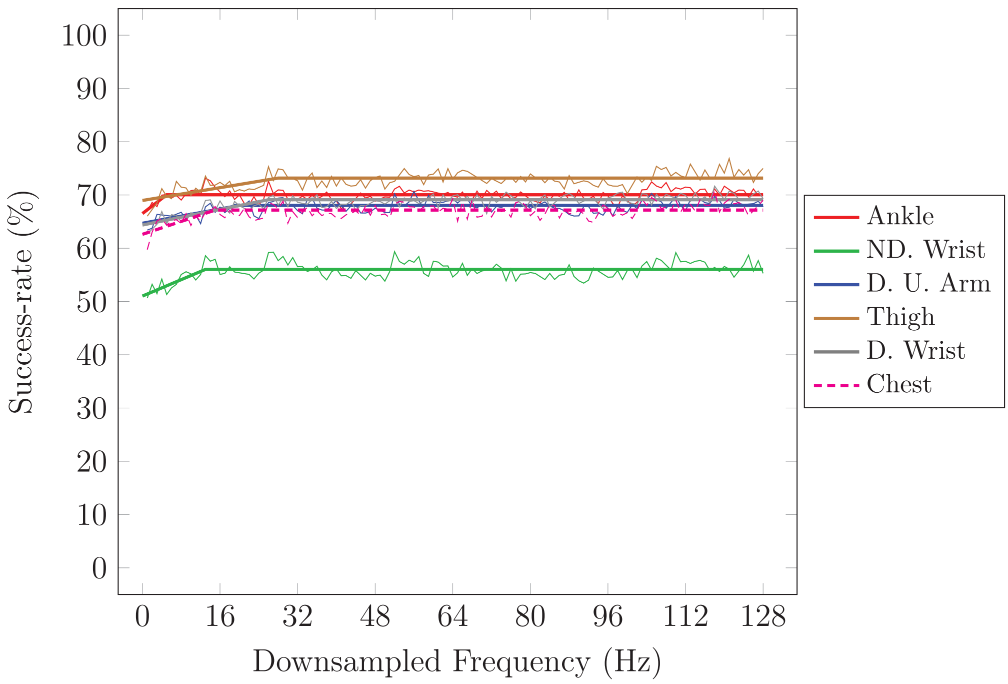
Washing Hands



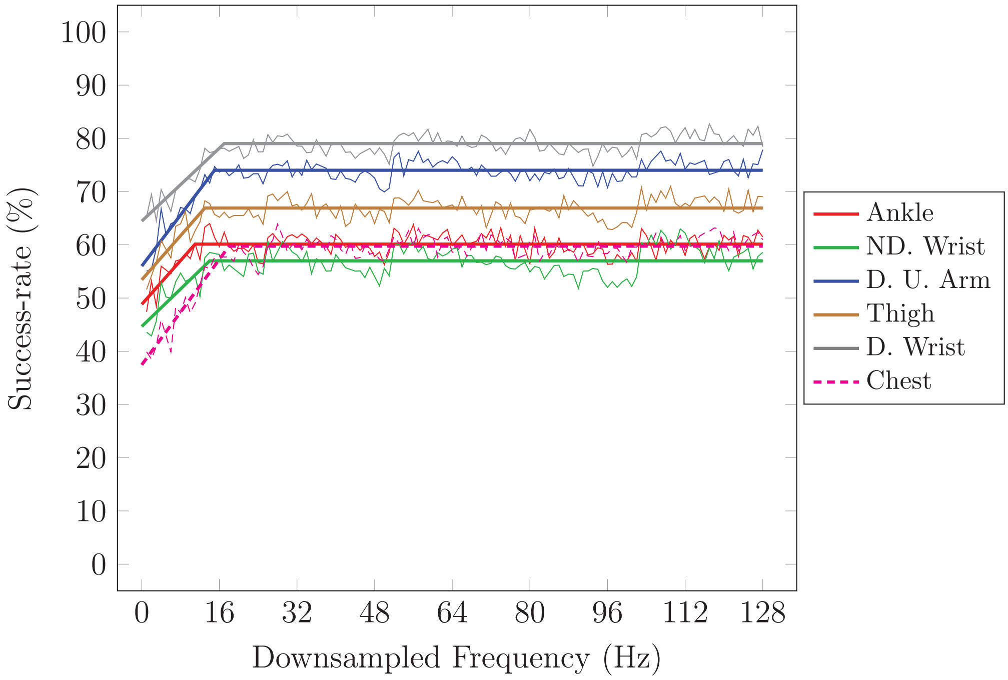
Washing Vegetables



Watching TV

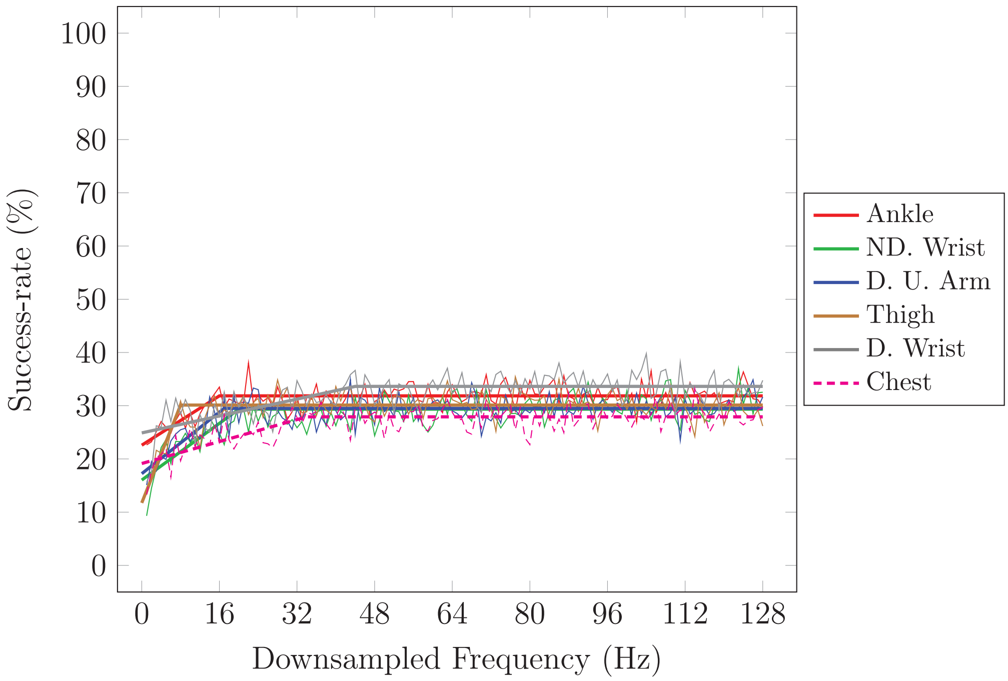


Writing

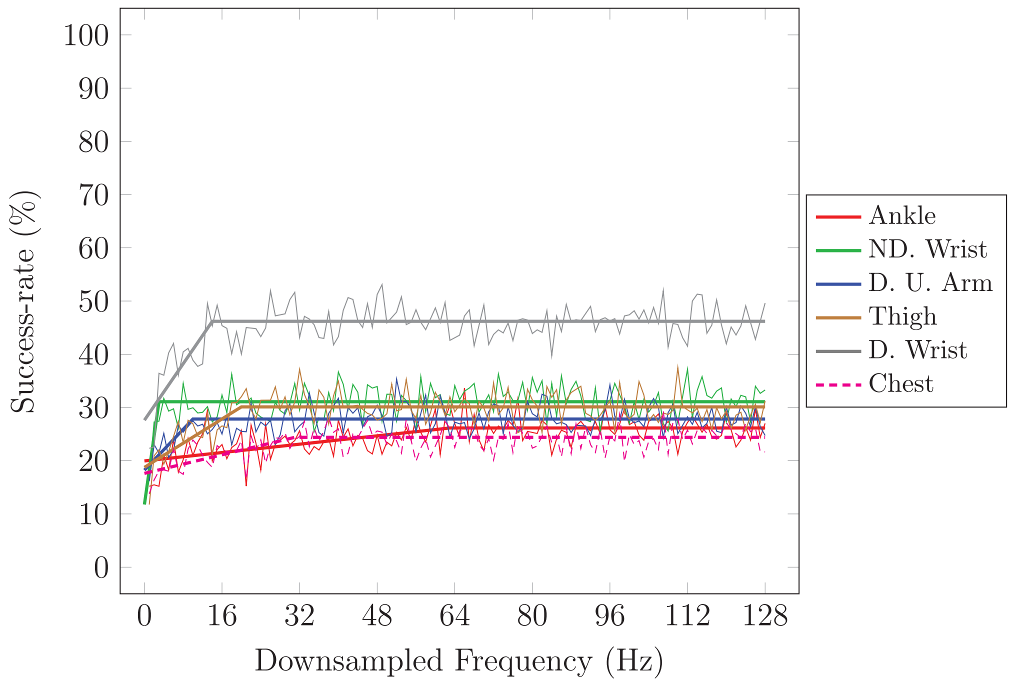


Kwapisz et al. (2011) - Gyroscope

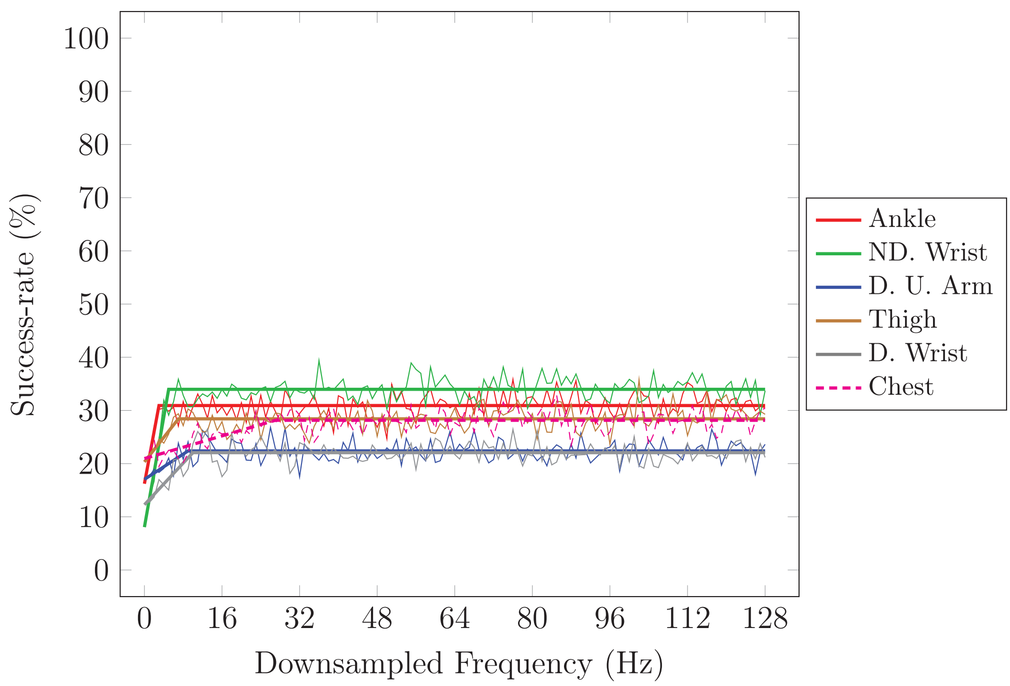
Brushing Teeth



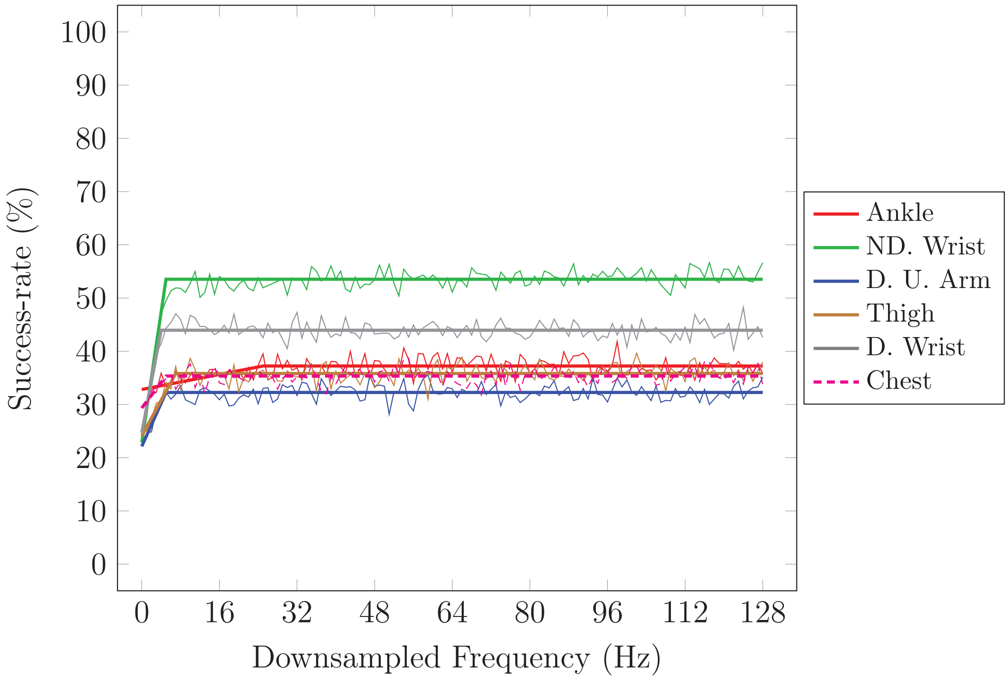
Dicing



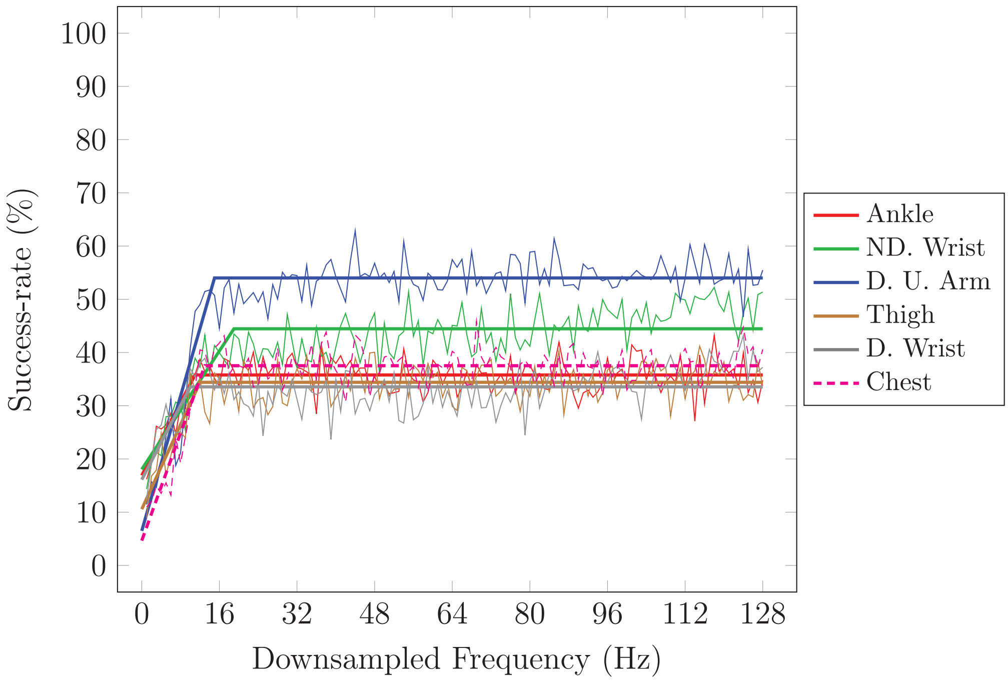
Dusting



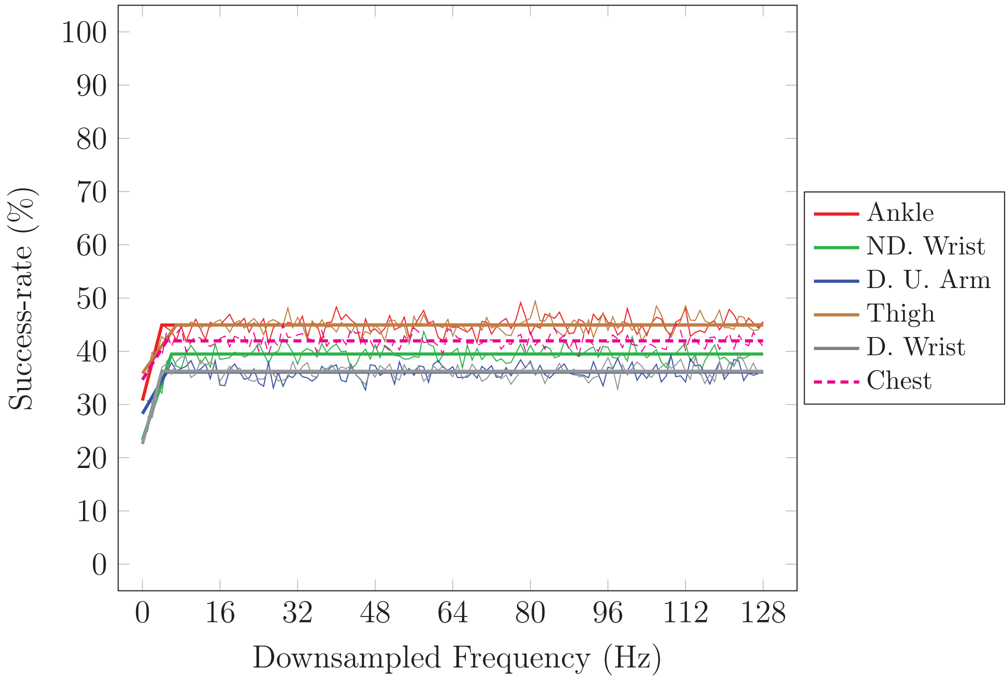
Folding Clothes



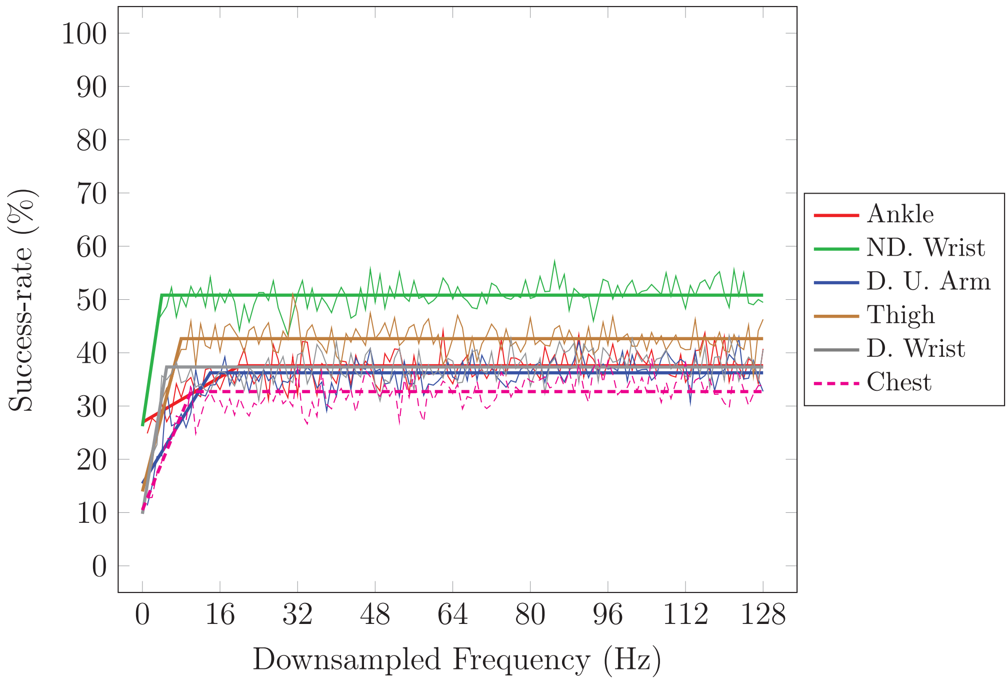
Grating



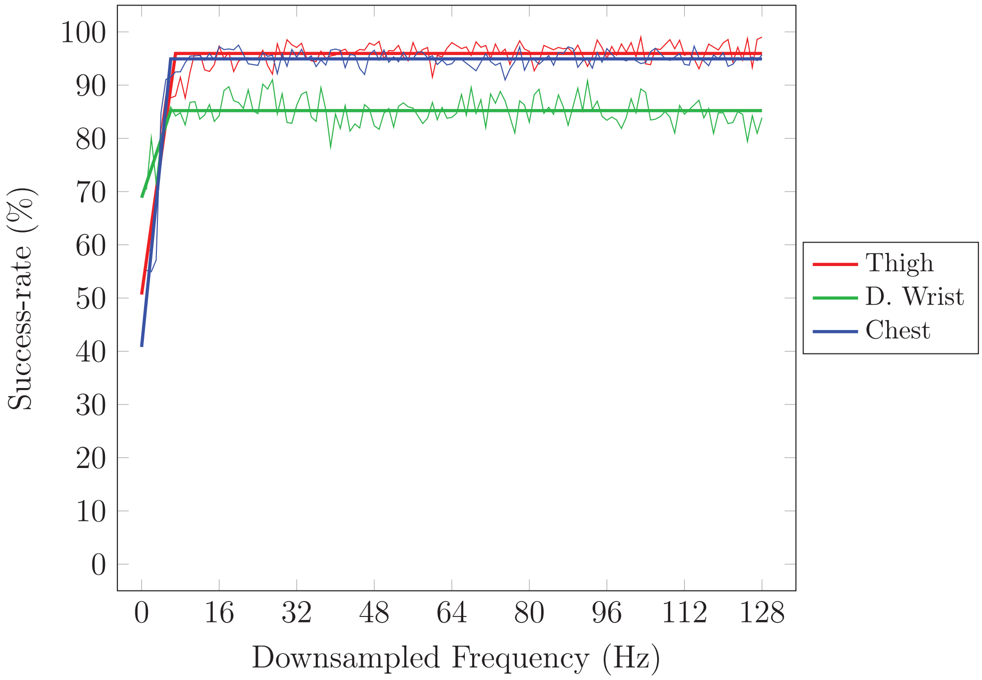
Ironing



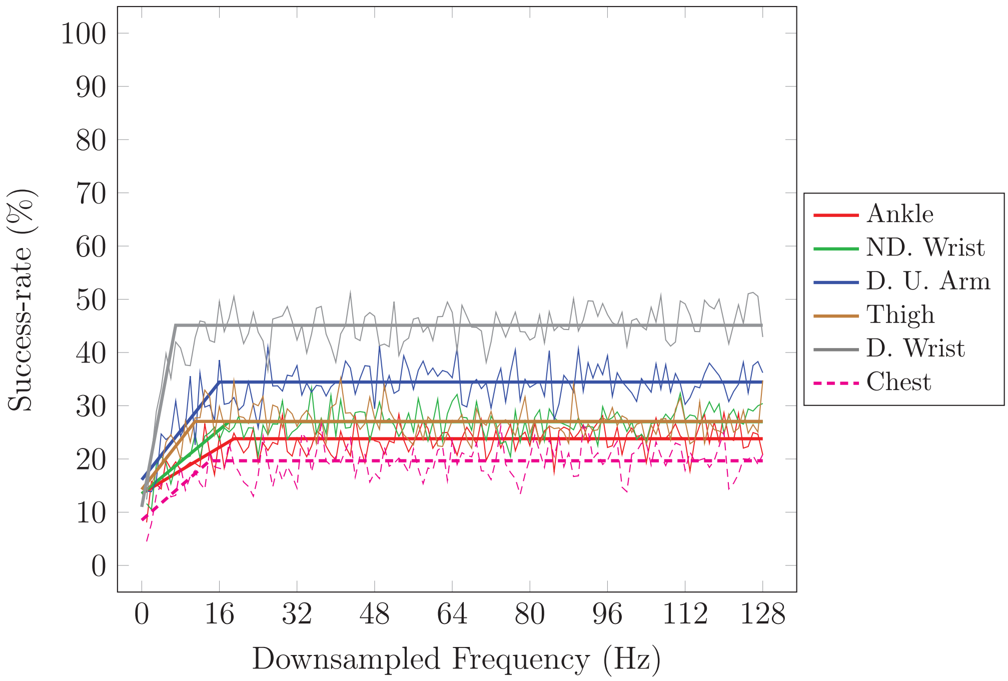
Peeling Vegetables



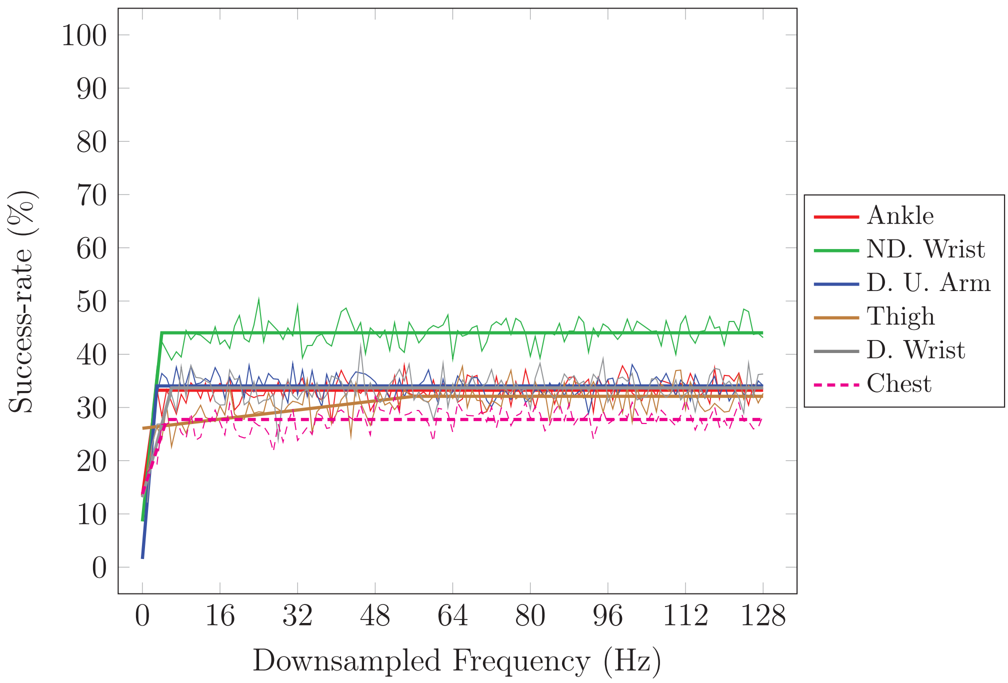
Running



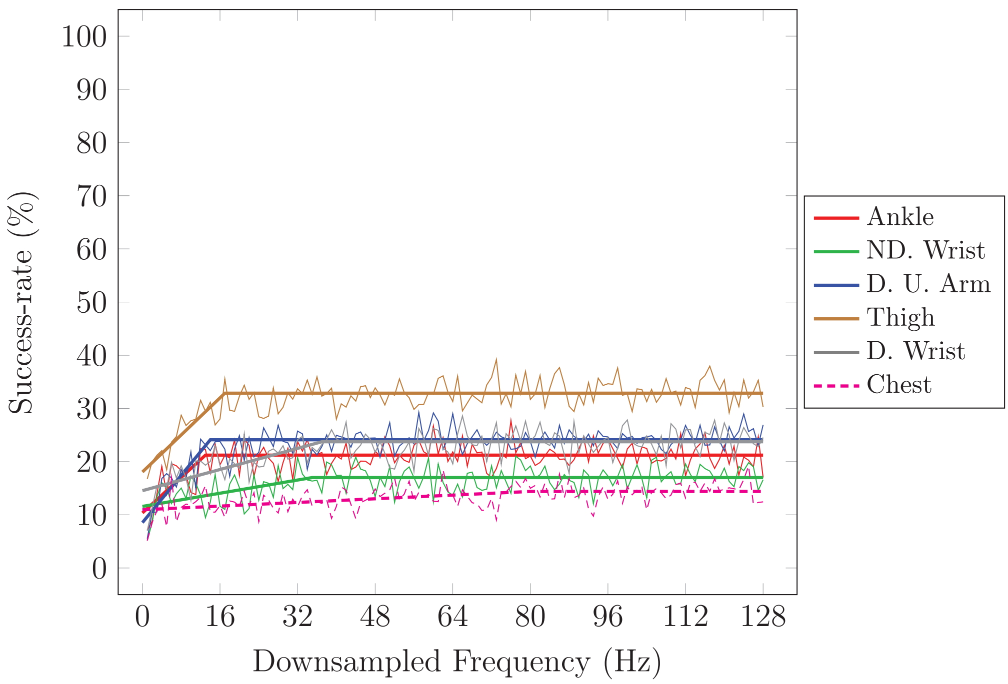
Stiring



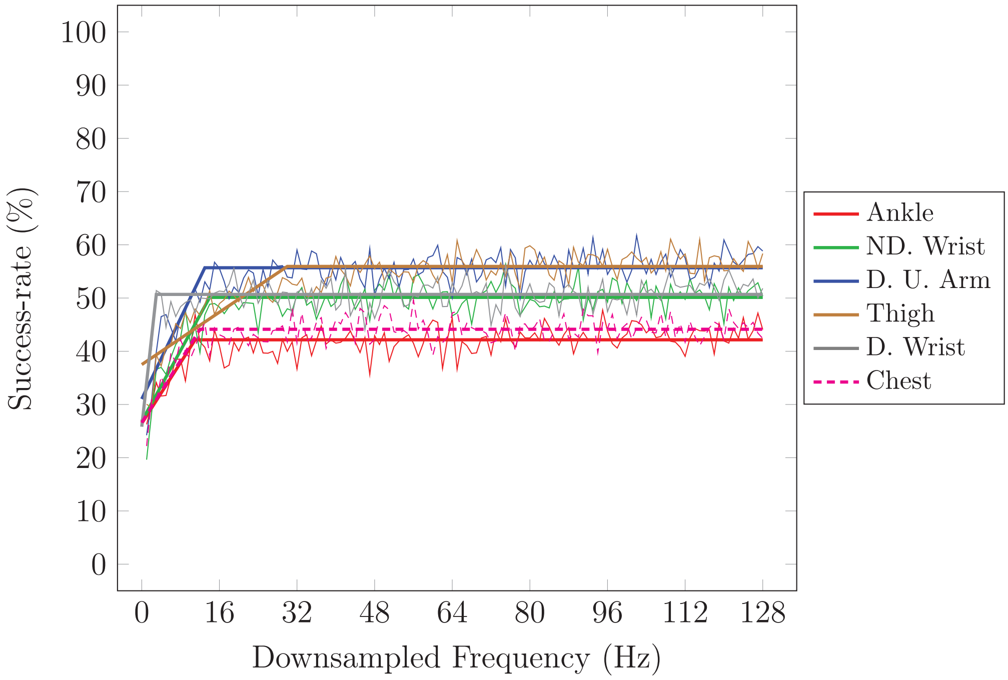
Sweeping



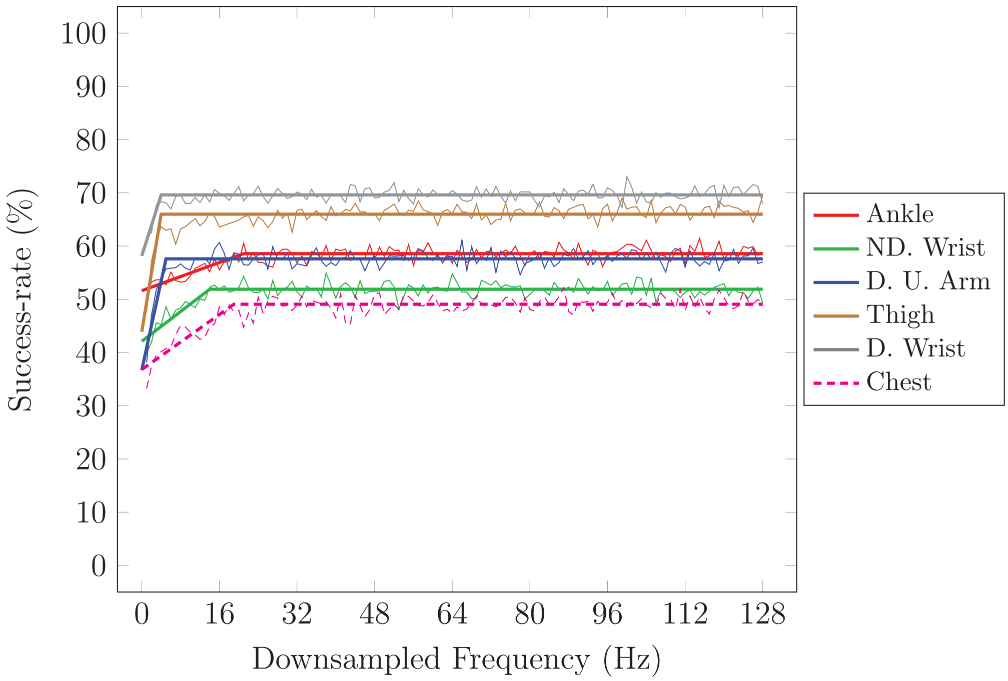
Talking on a Phone



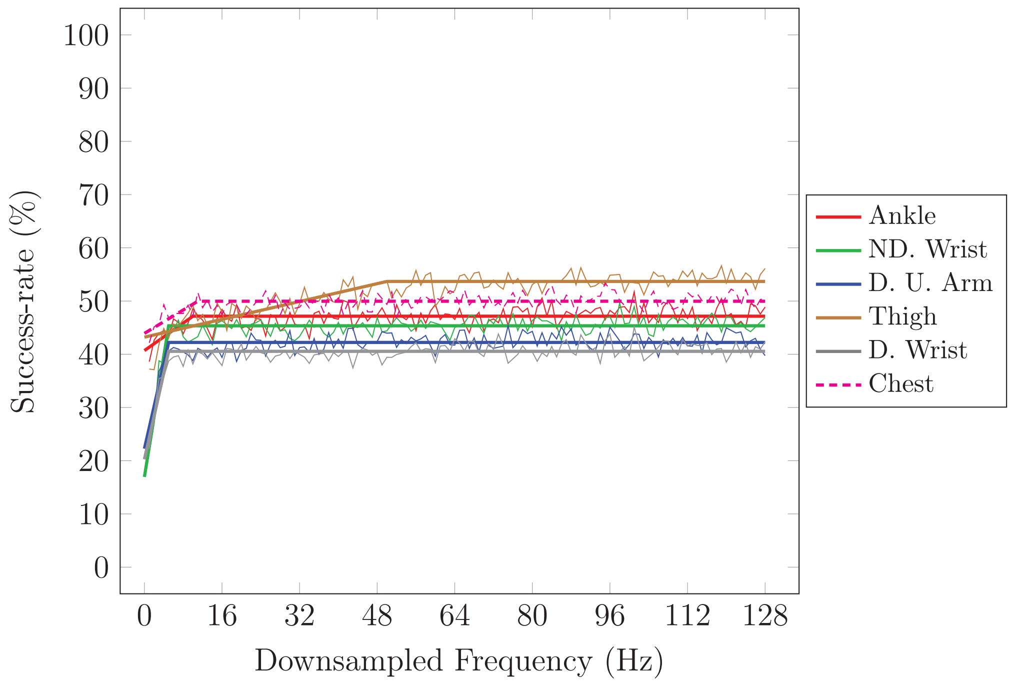
Texting on a Phone



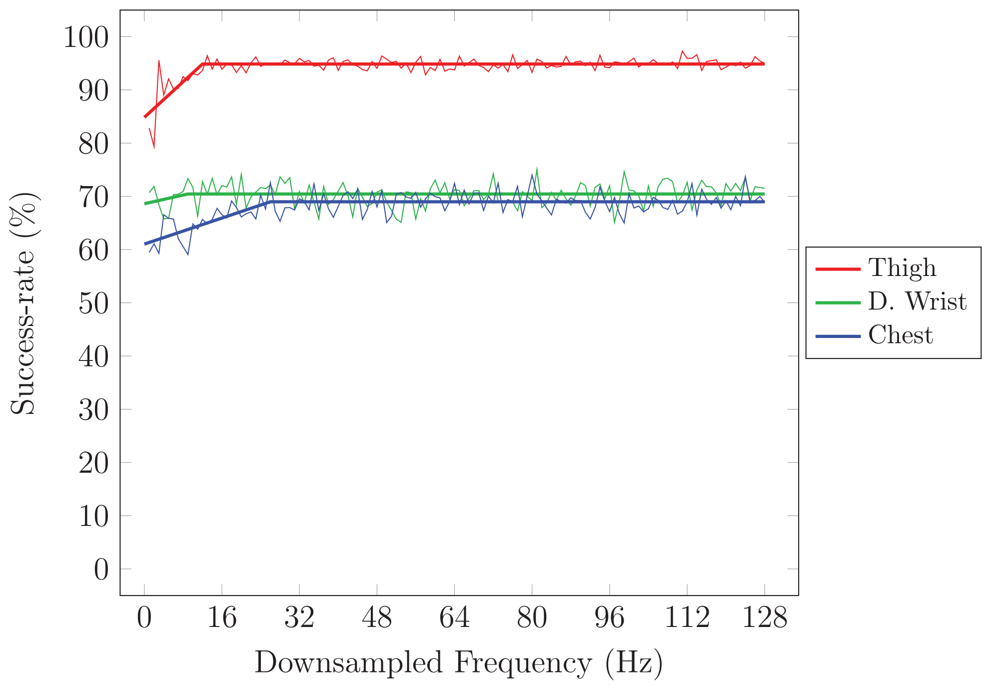
Using a PC



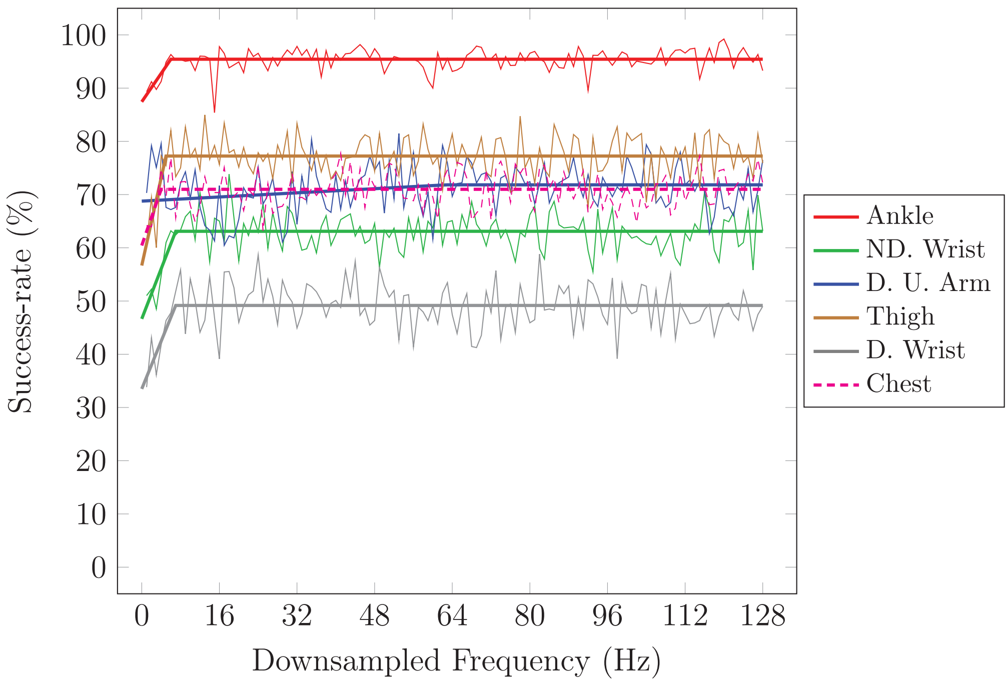
Vacuuming



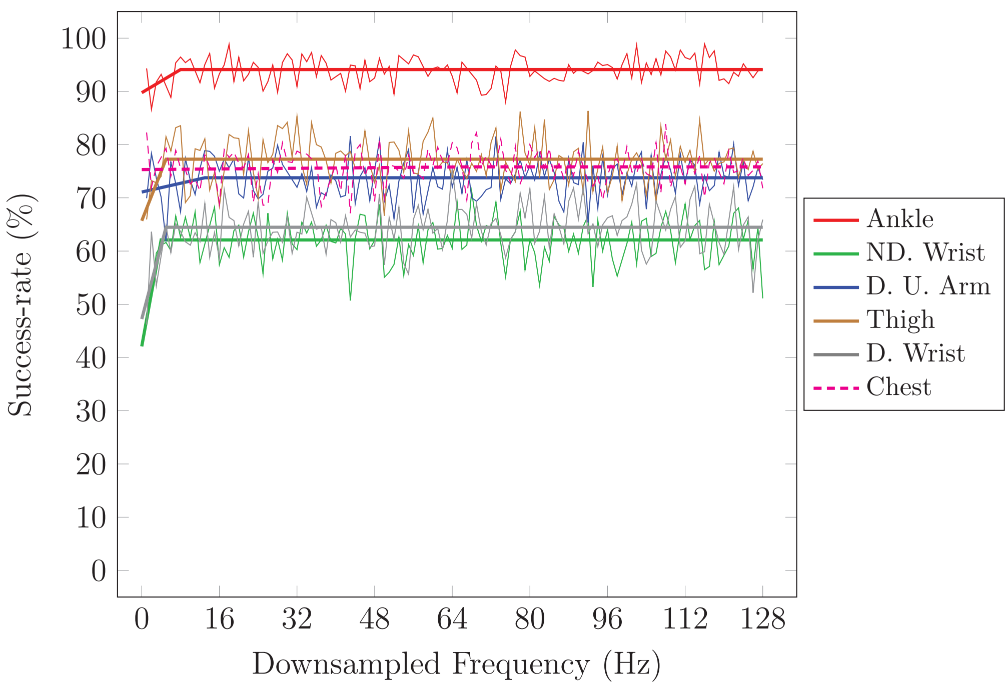
Walking (Flat Ground)



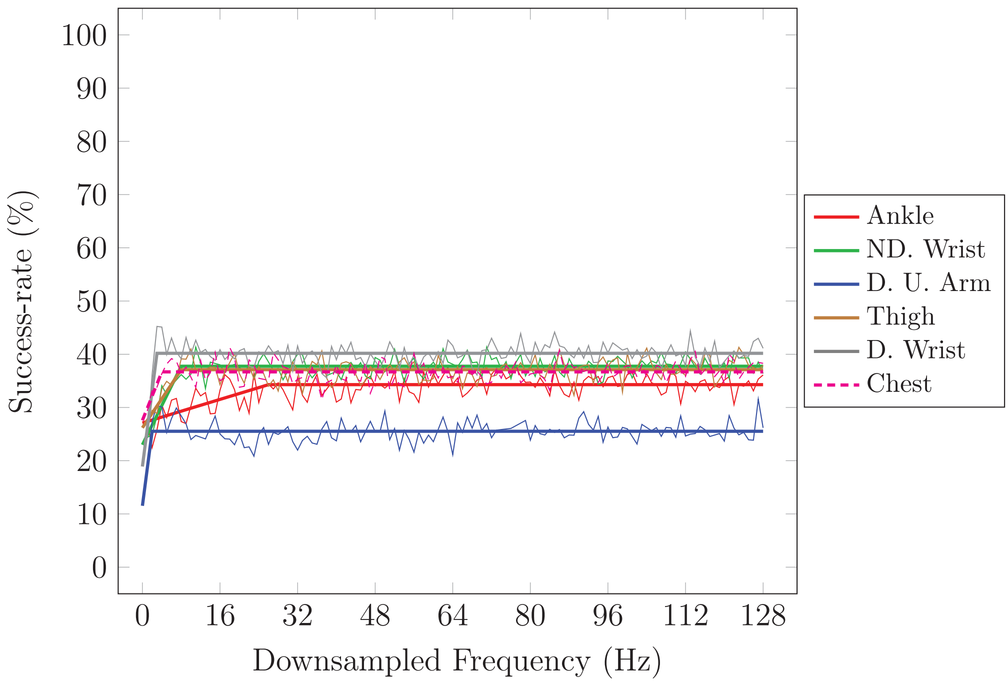
Walking Down Stairs



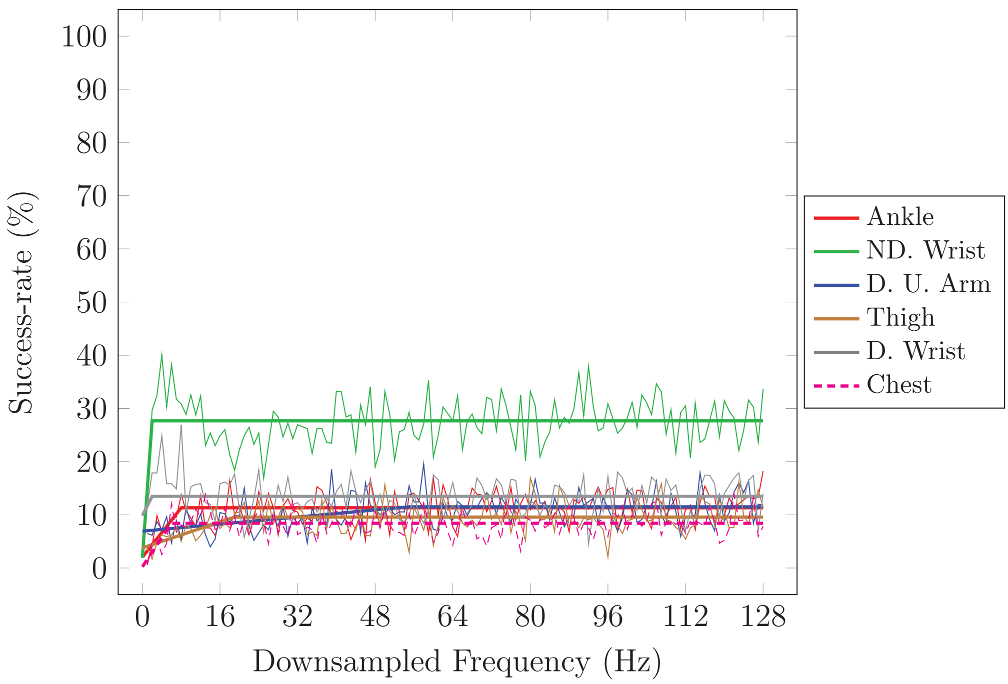
Walking Up Stairs



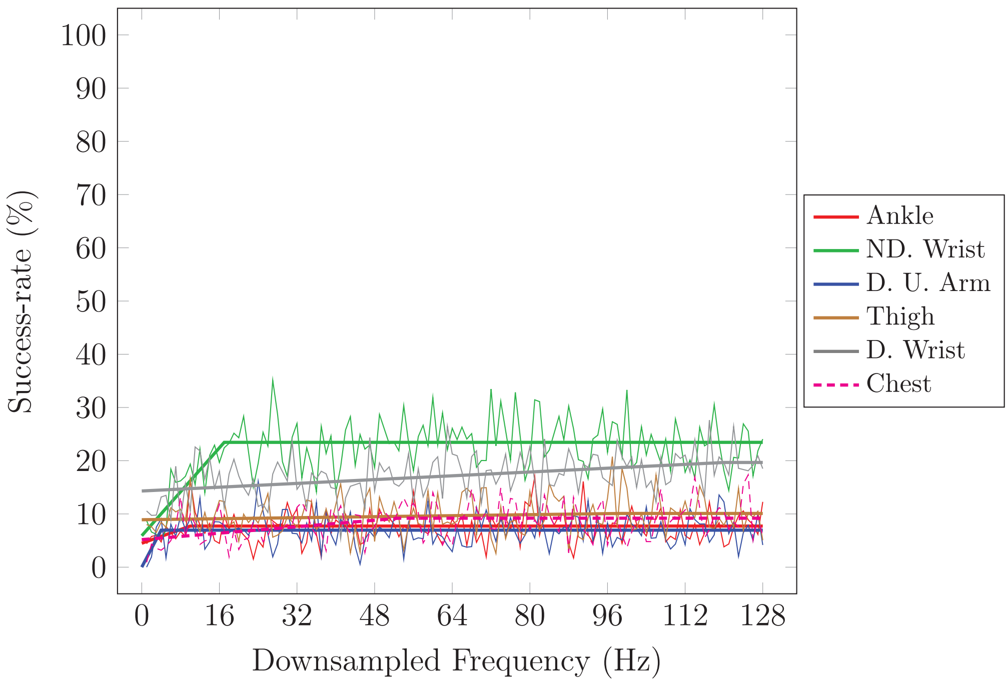
Washing Dishes



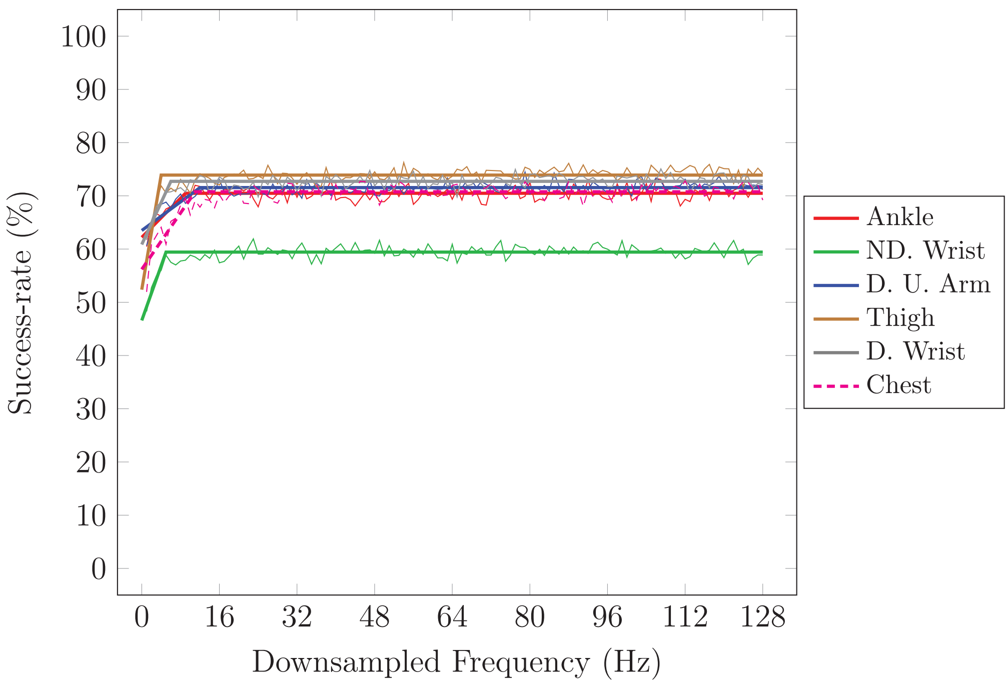
Washing Hands



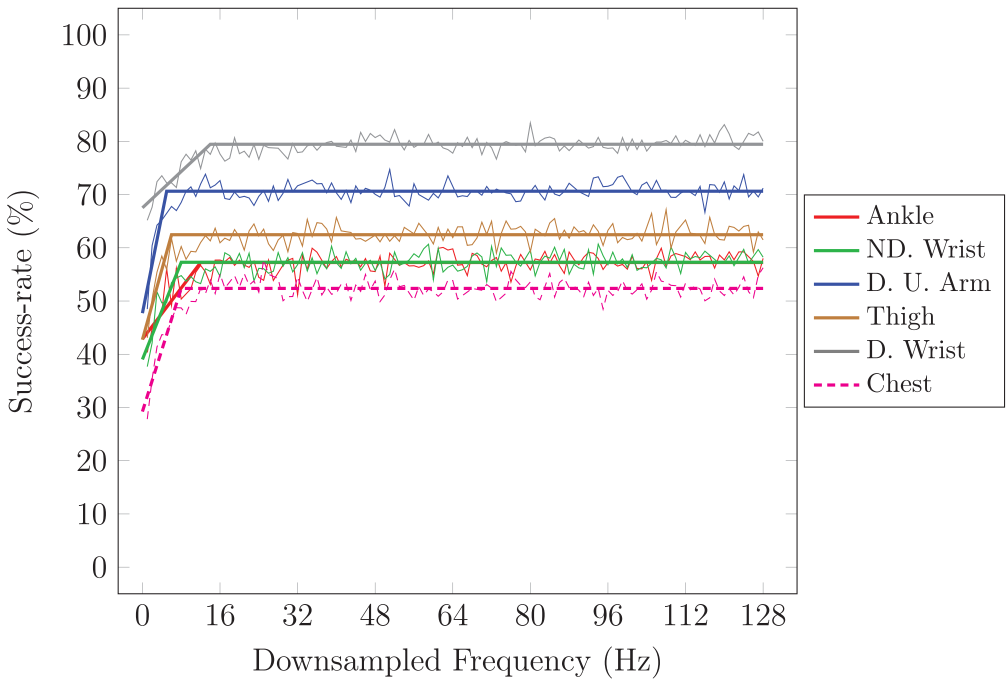
Washing Vegetables



Watching TV

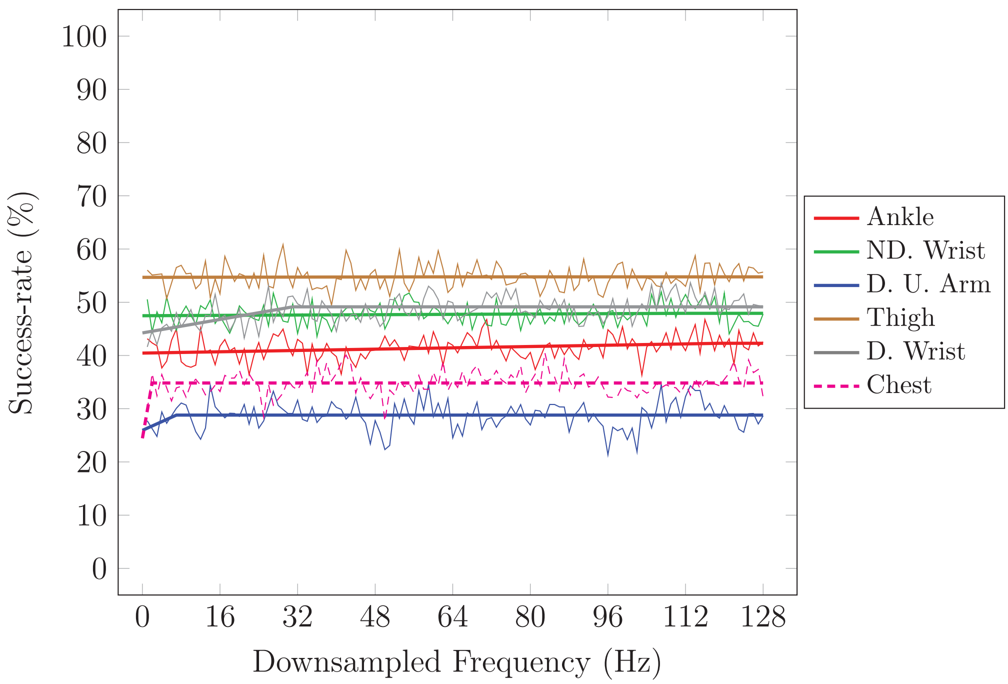


Writing

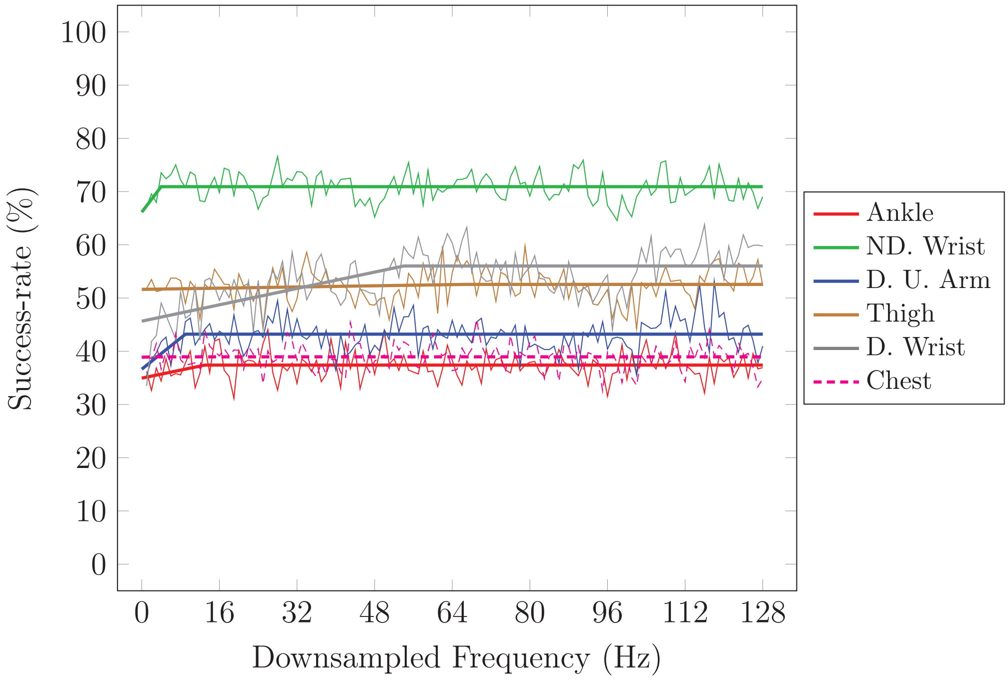


Bao and Intille (2004) - Orientation

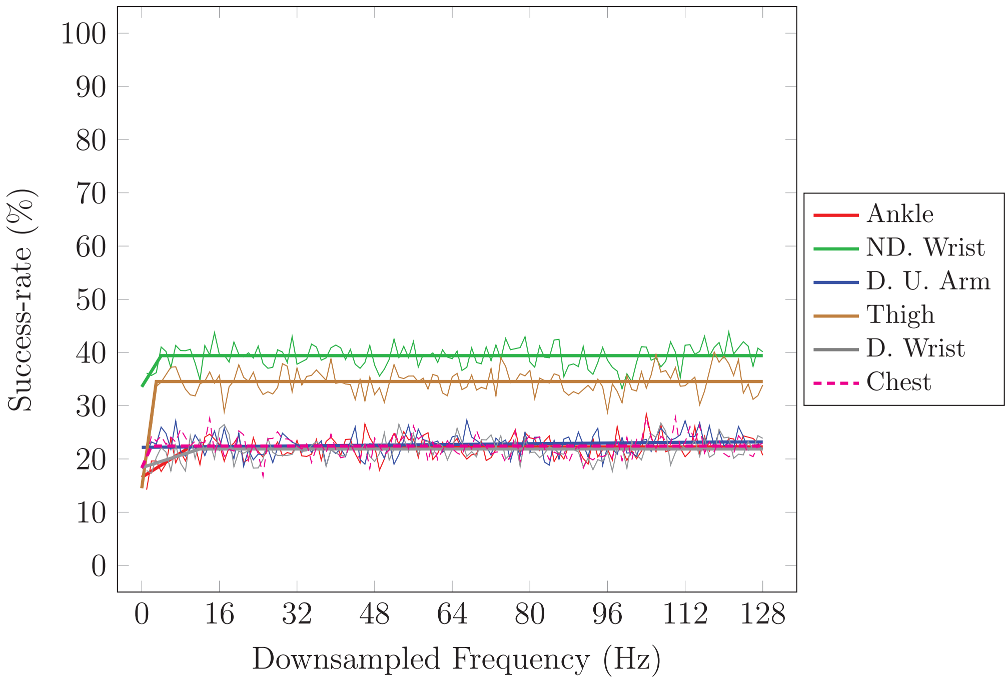
Brushing Teeth



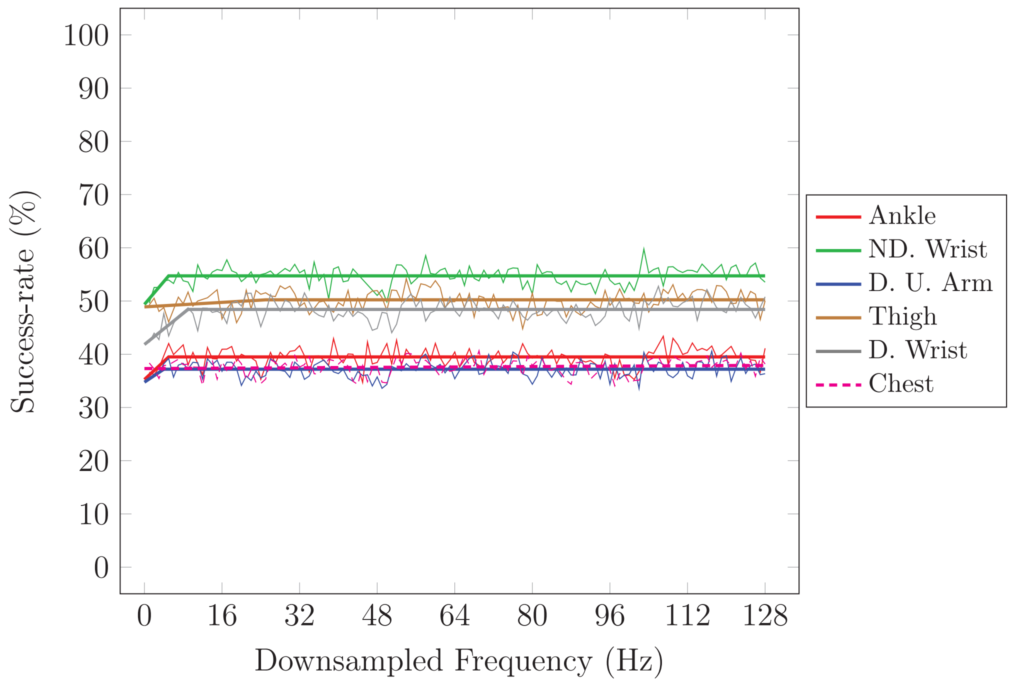
Dicing



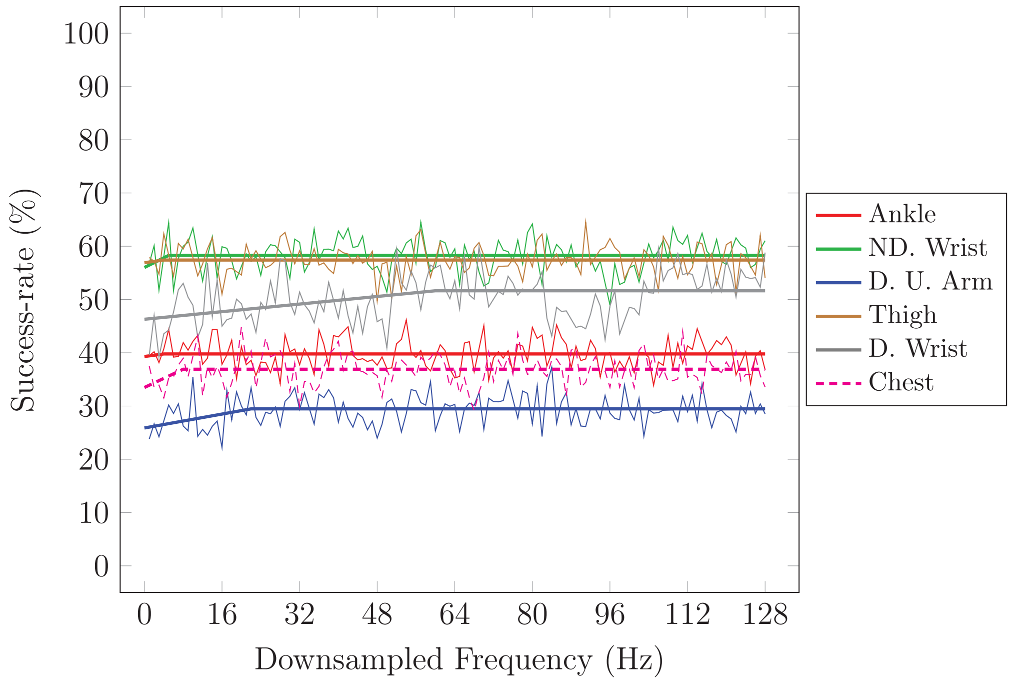
Dusting



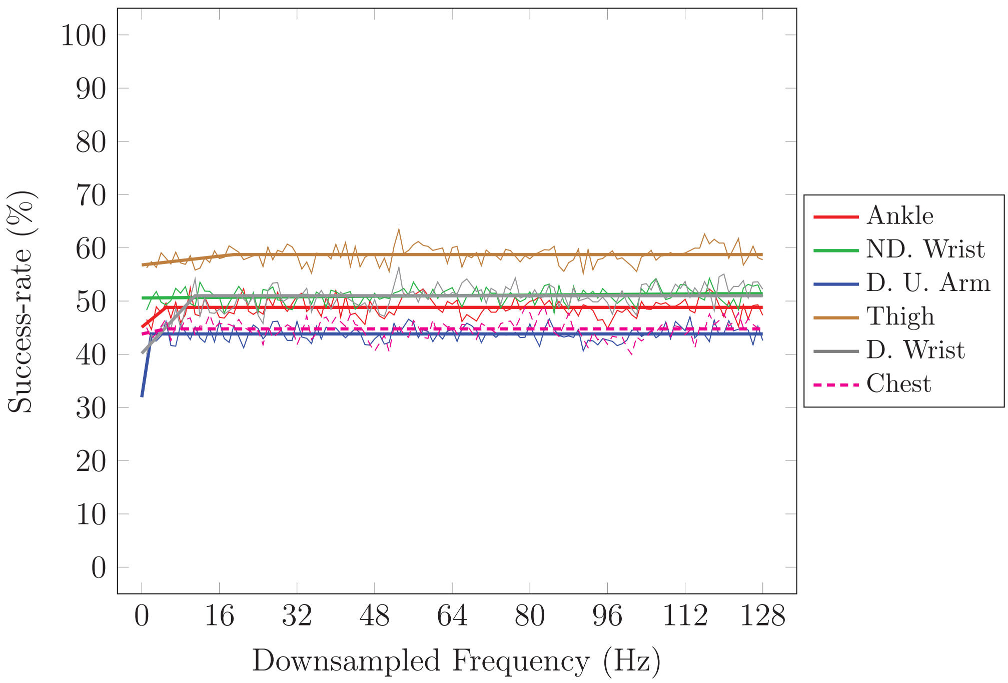
Folding Clothes



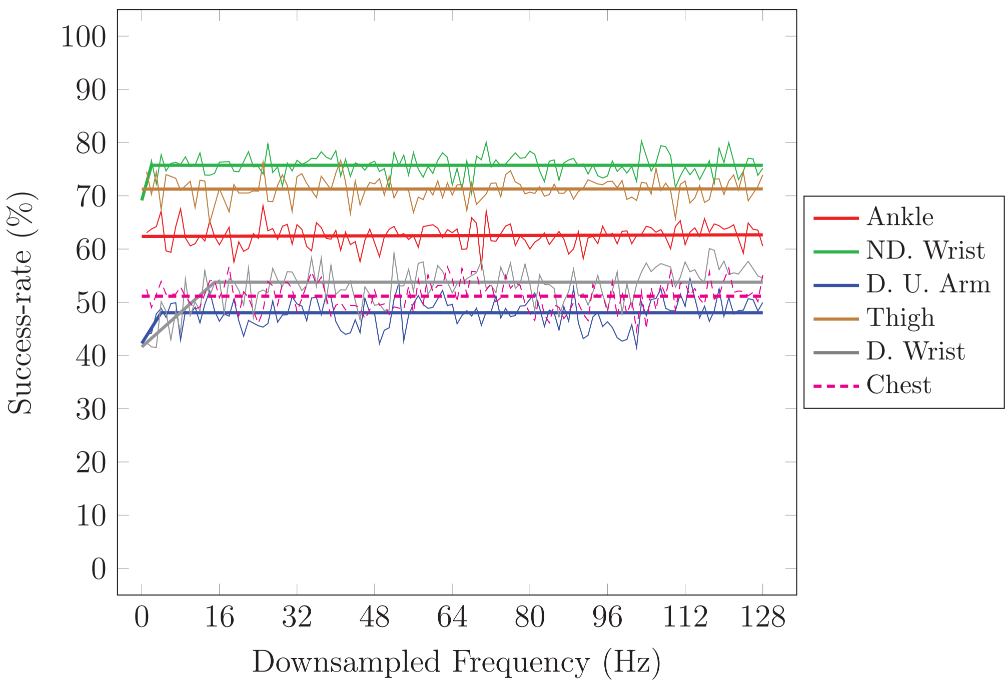
Grating



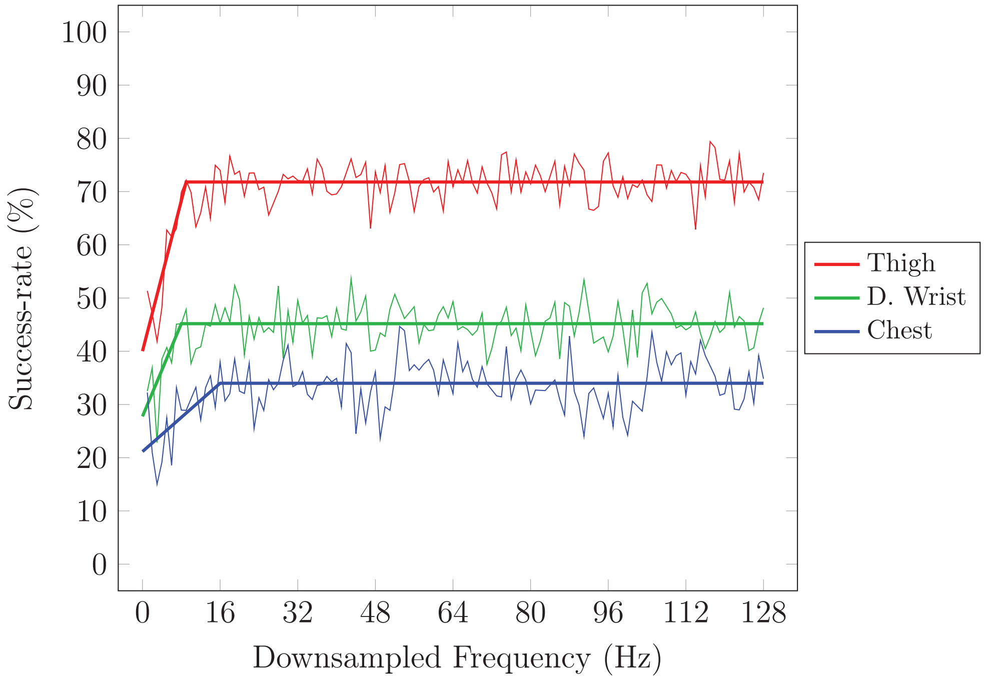
Ironing



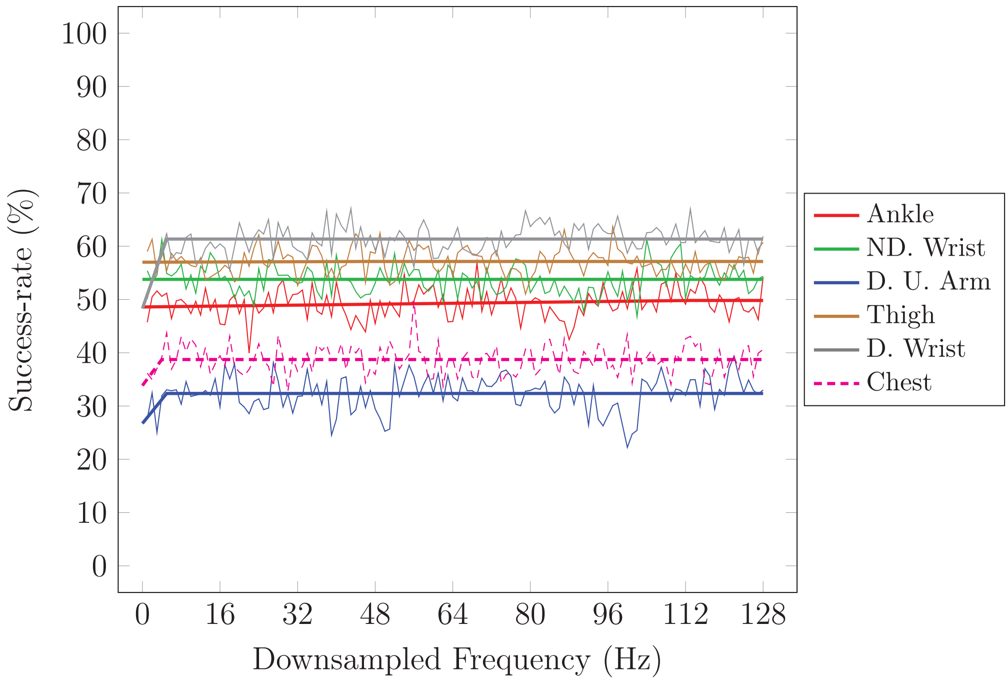
Peeling Vegetables



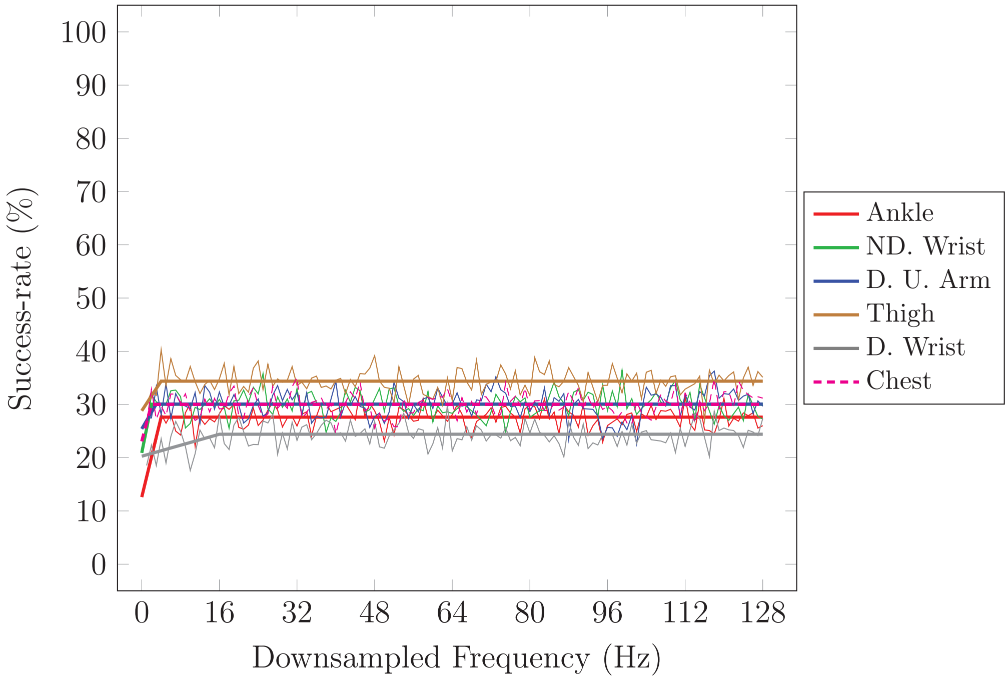
Running



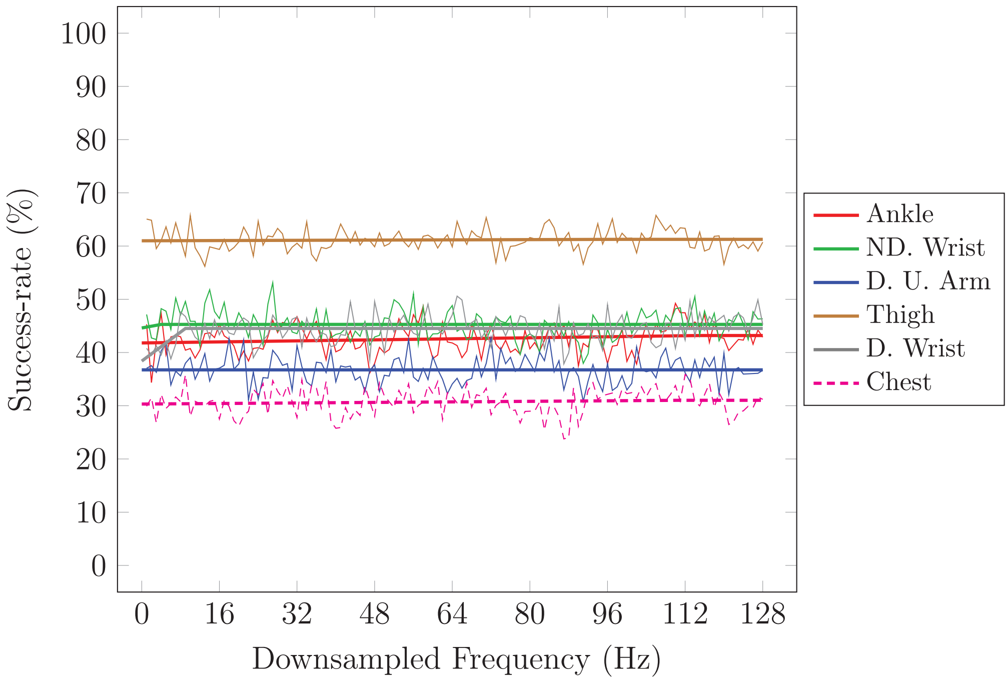
Stiring



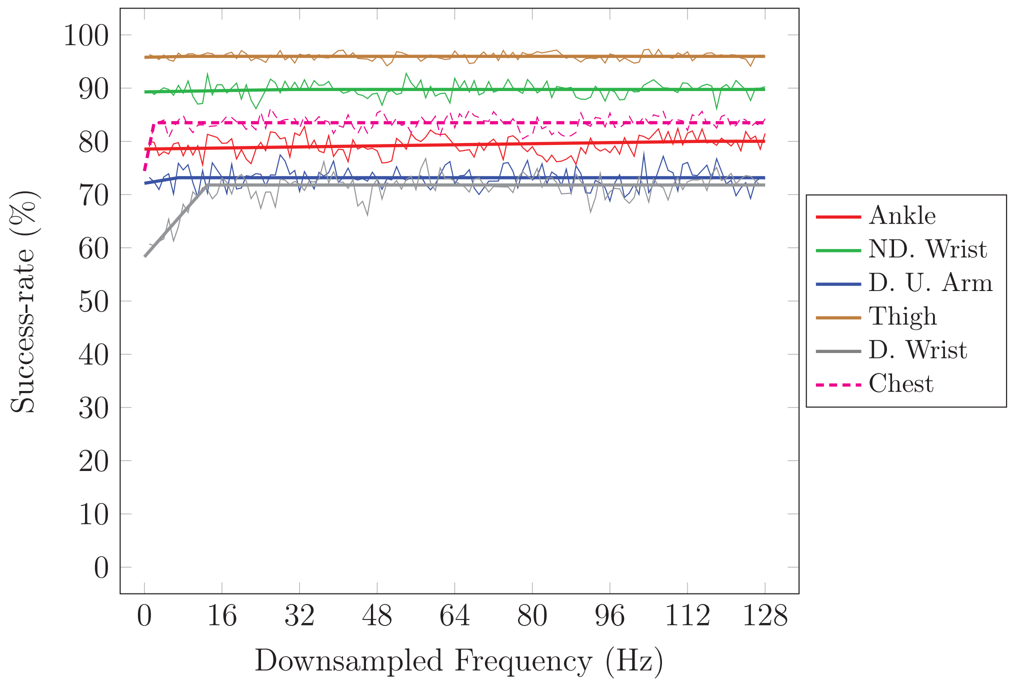
Sweeping



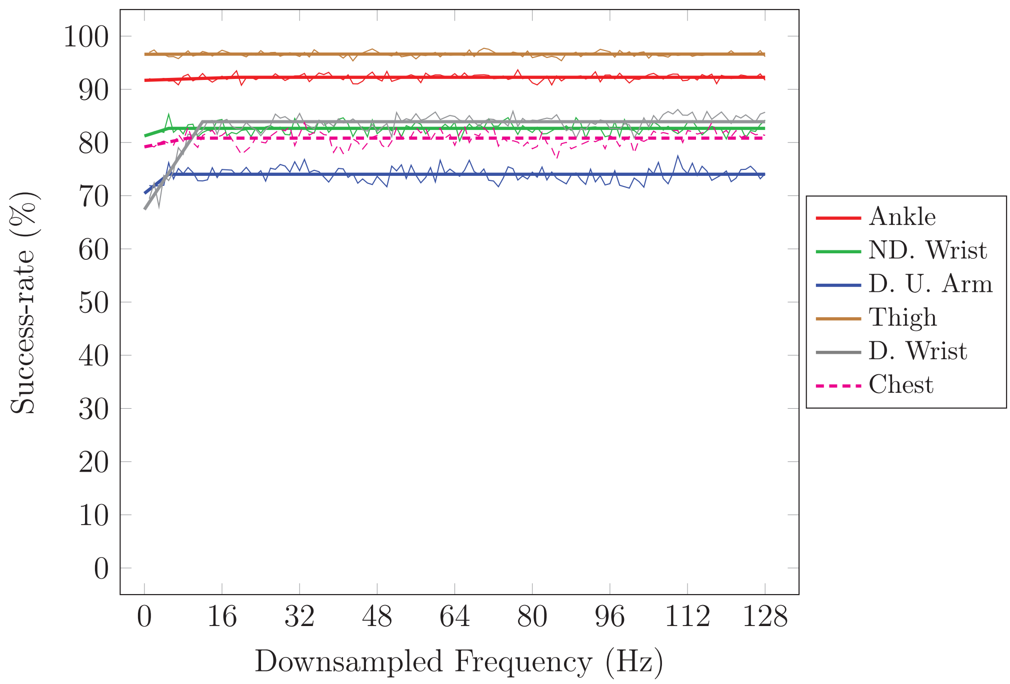
Talking on a Phone



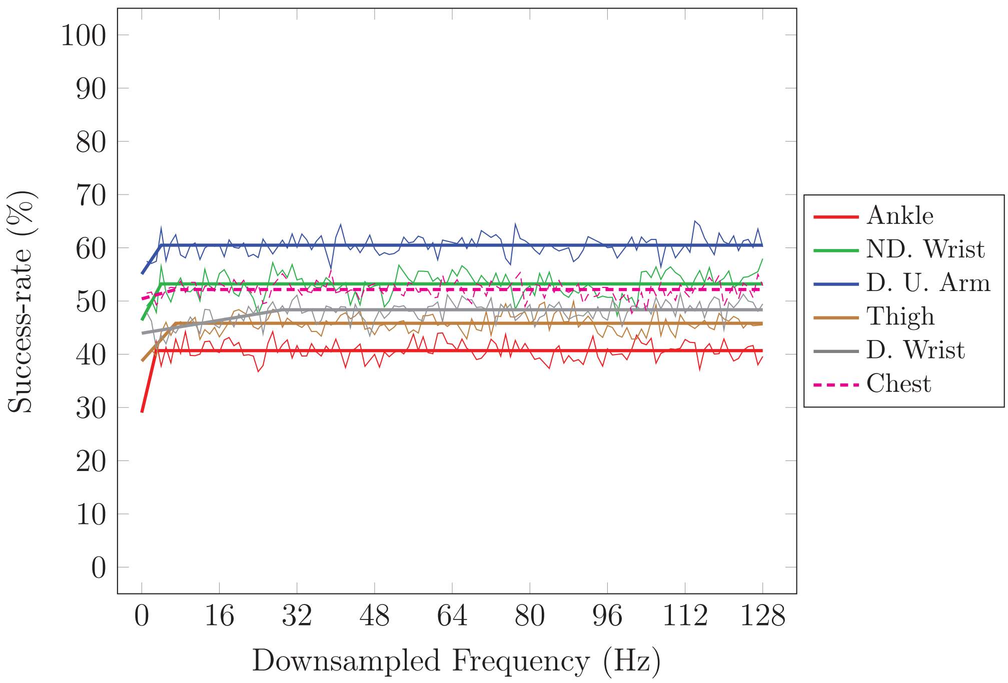
Texting on a Phone



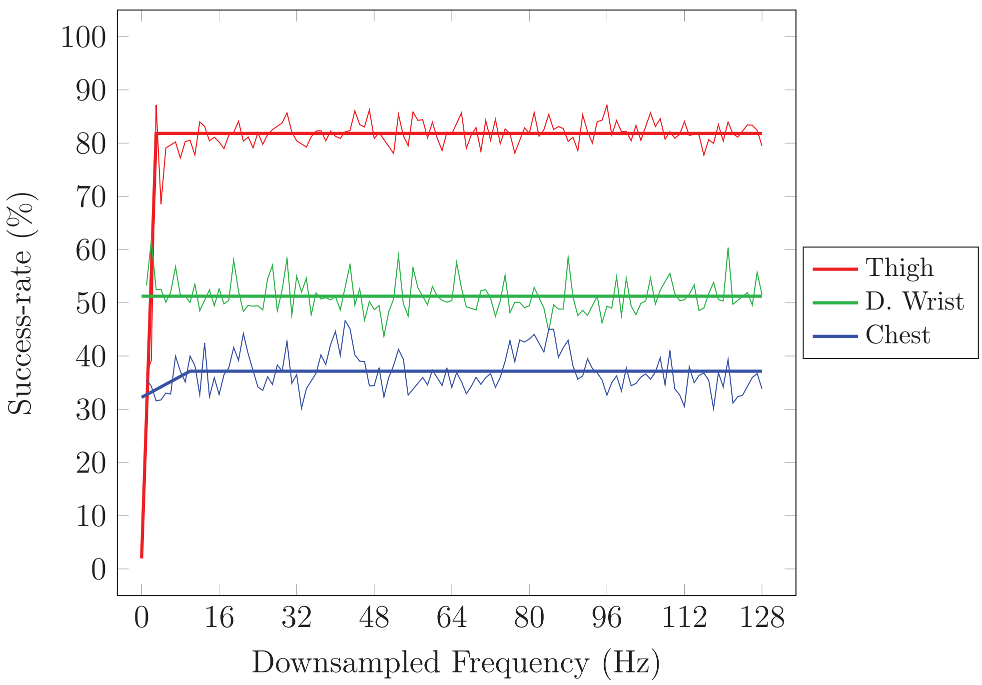
Using a PC



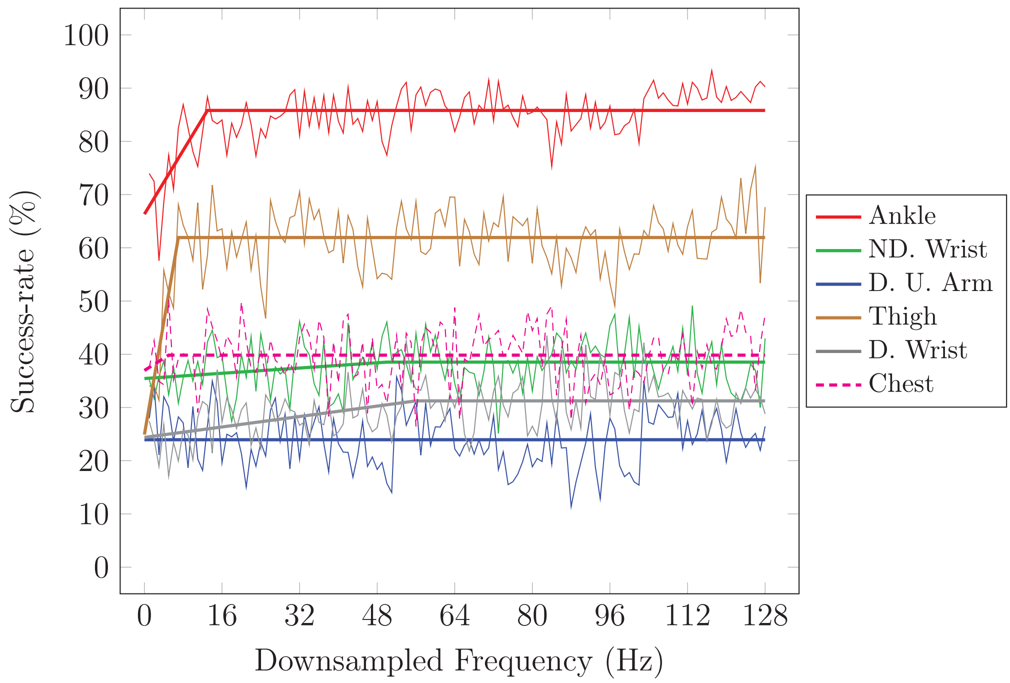
Vacuuming



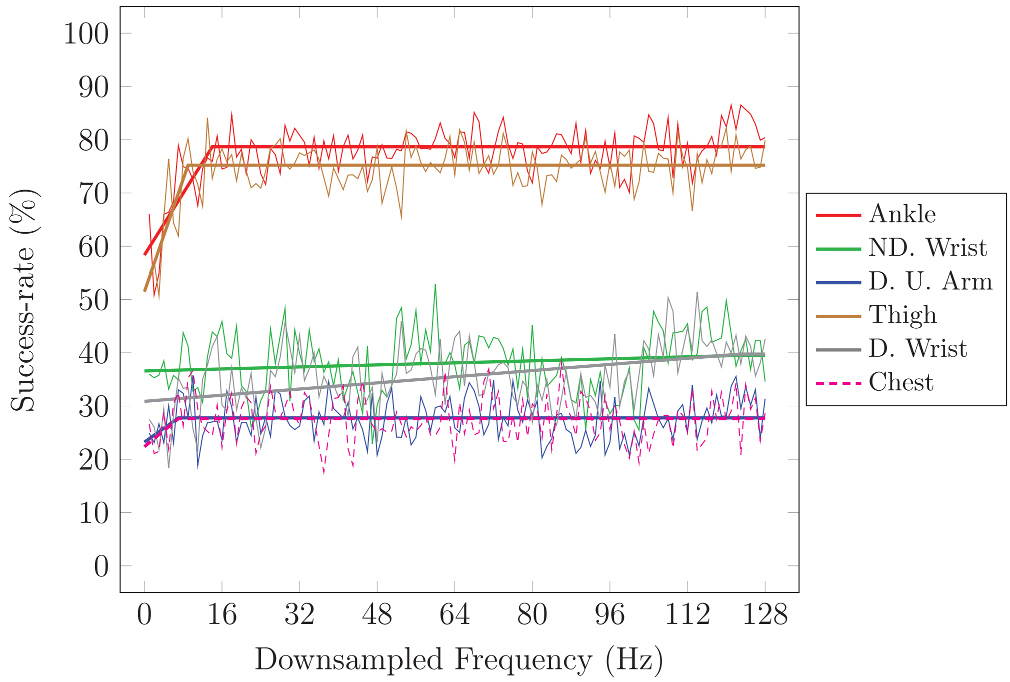
Walking (Flat Ground)



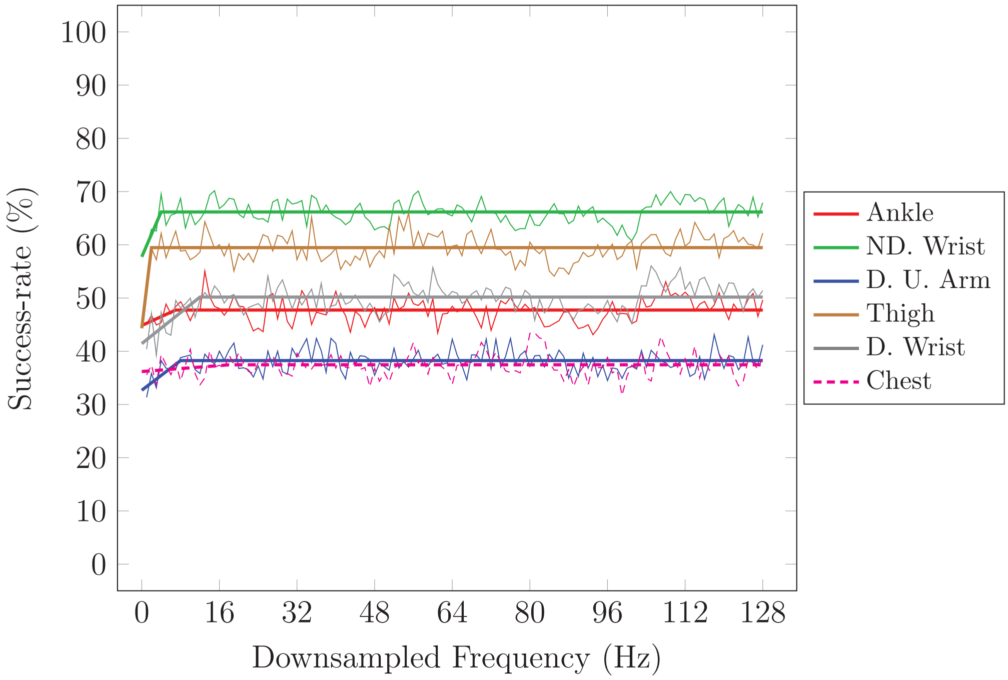
Walking Down Stairs



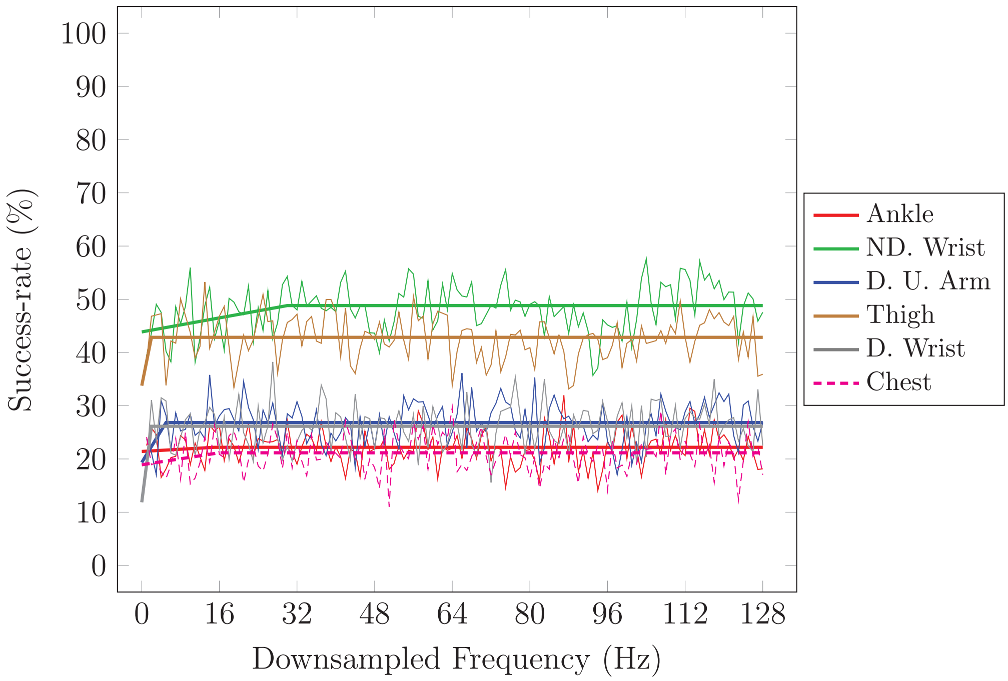
Walking Up Stairs



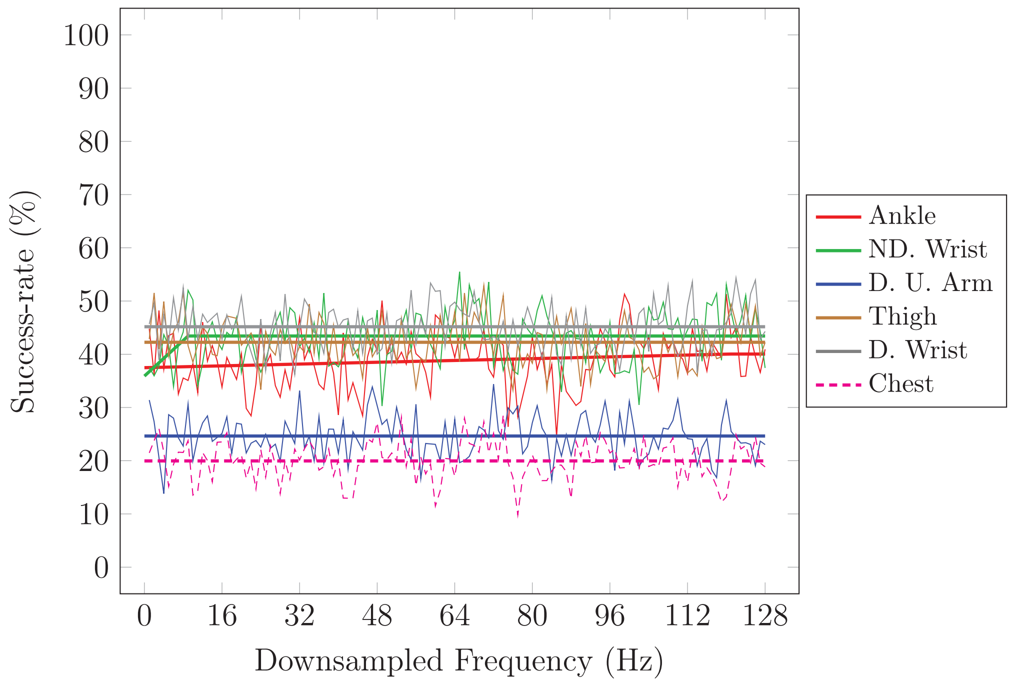
Washing Dishes



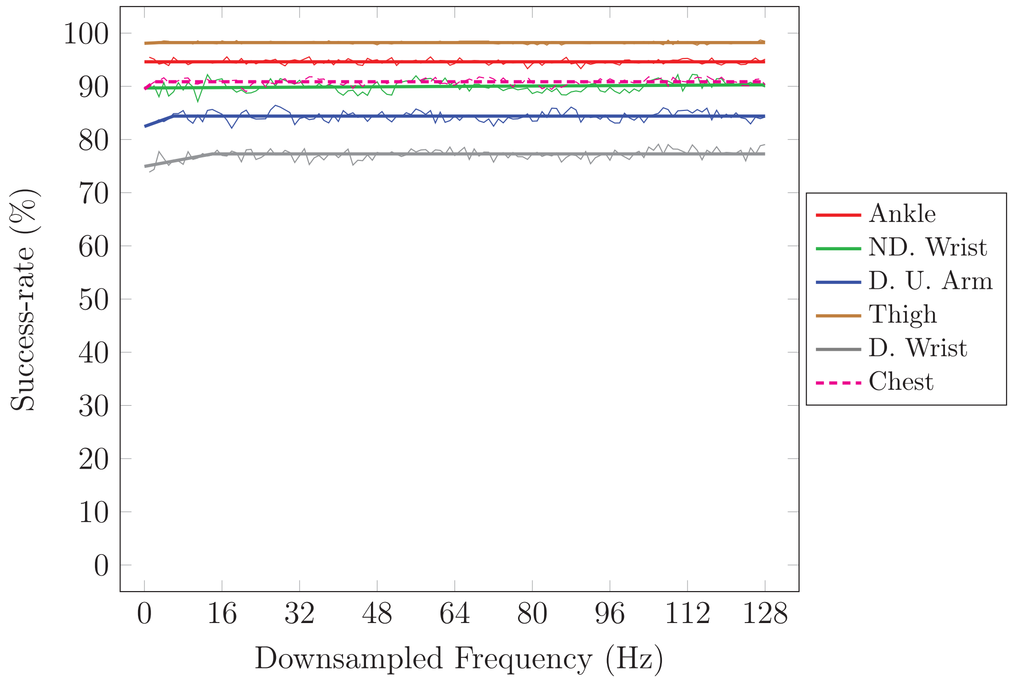
Washing Hands



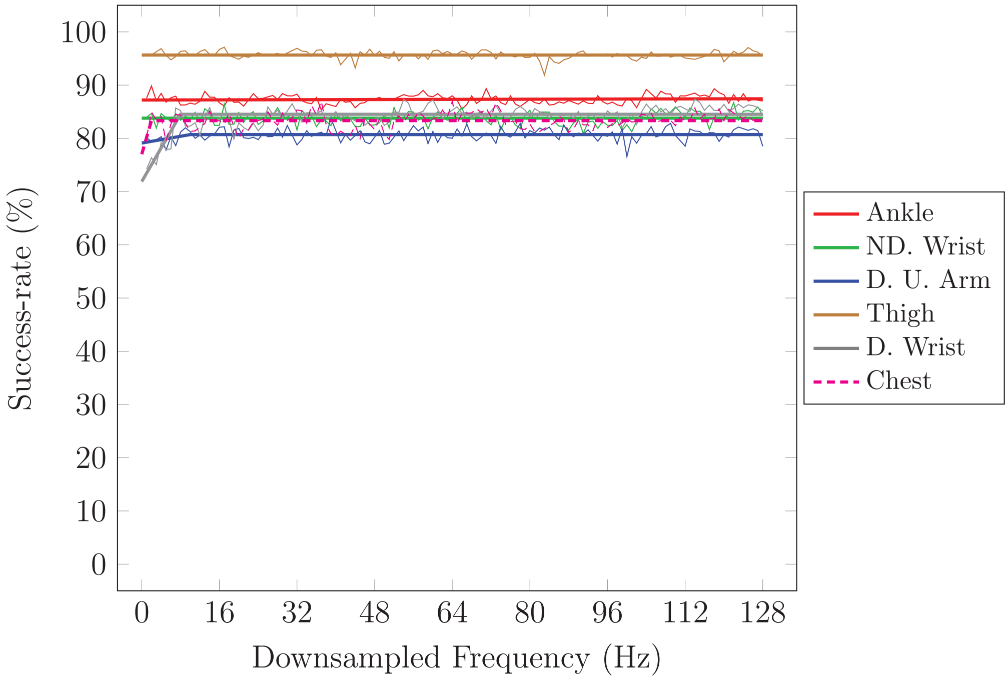
Washing Vegetables



Watching TV

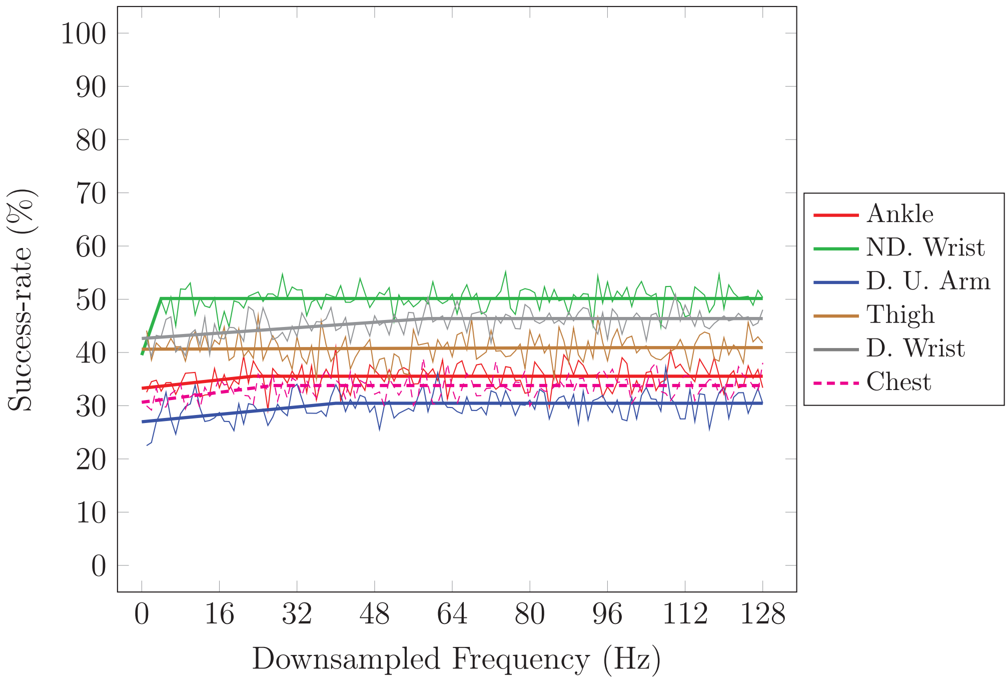


Writing

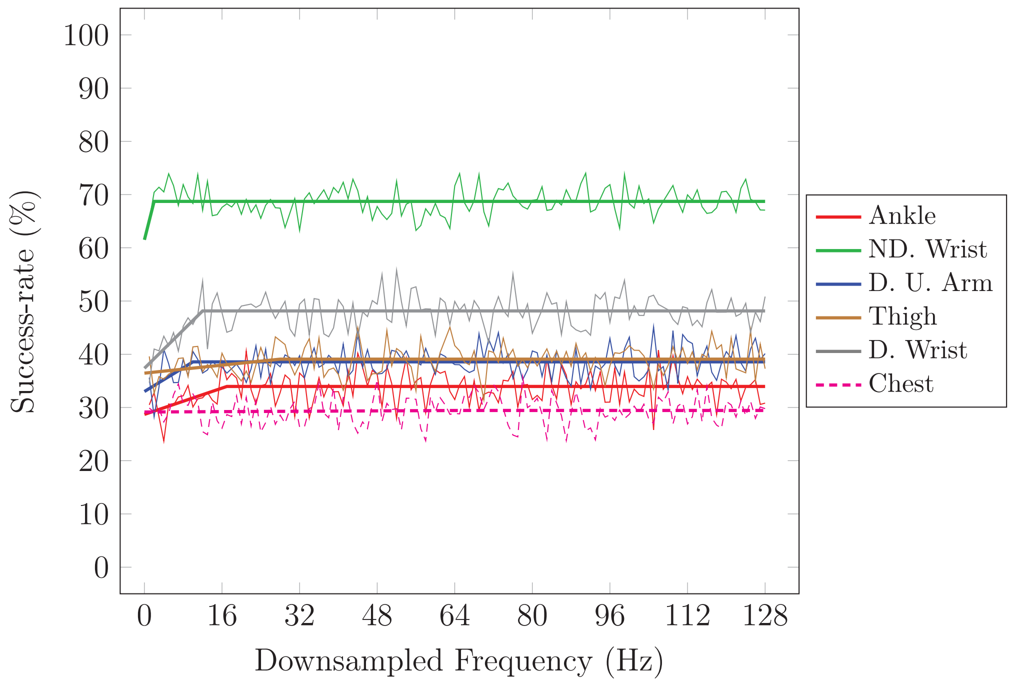


Kwapisz et al. (2011) - Orientation

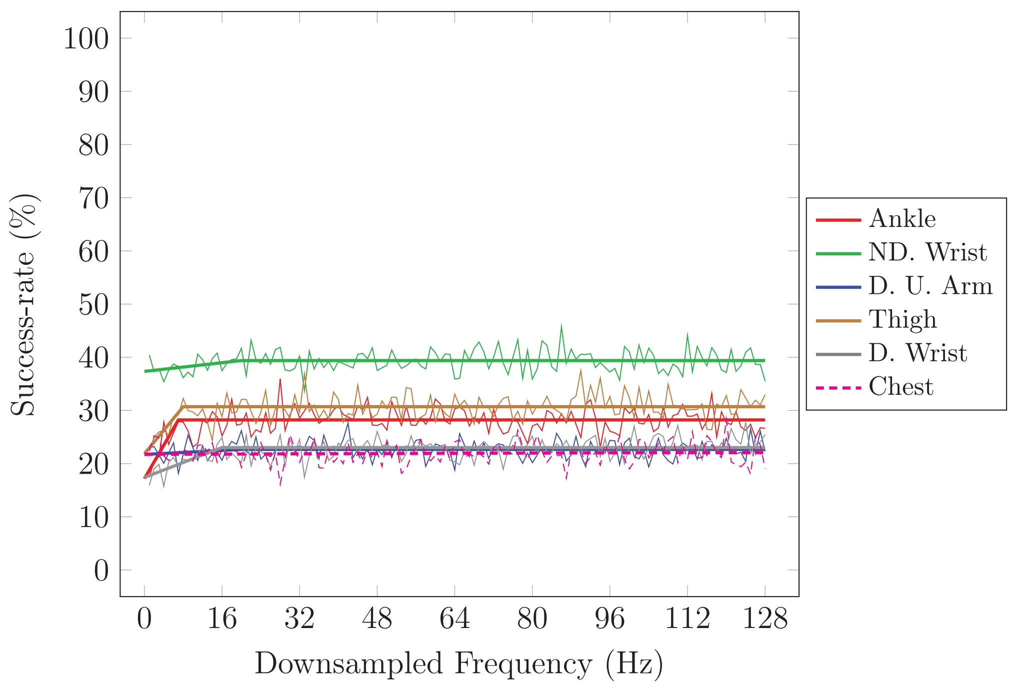
Brushing Teeth



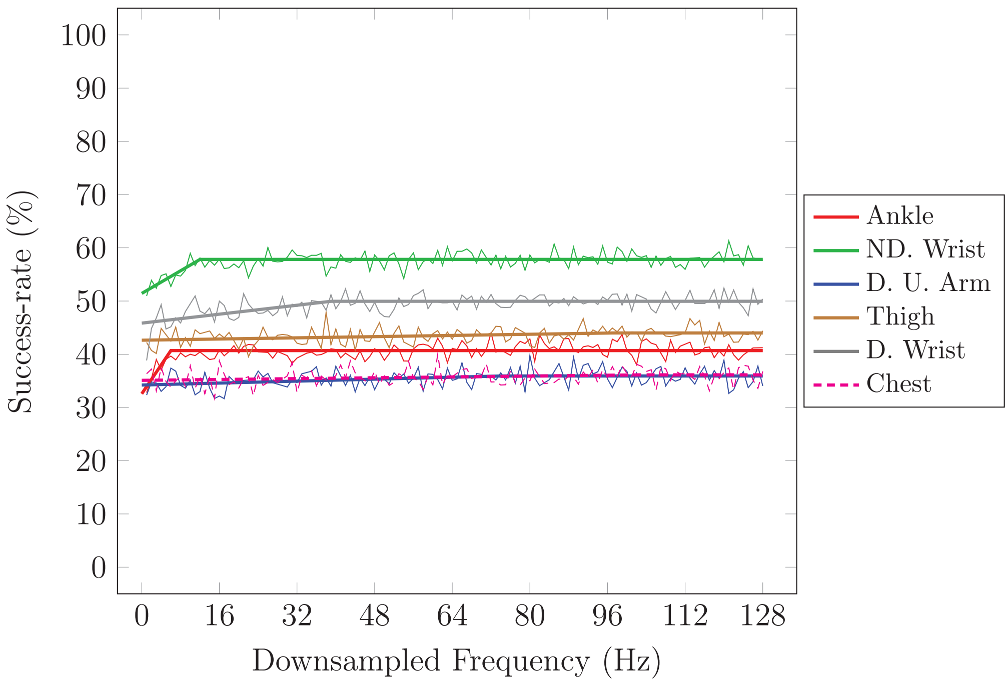
Dicing



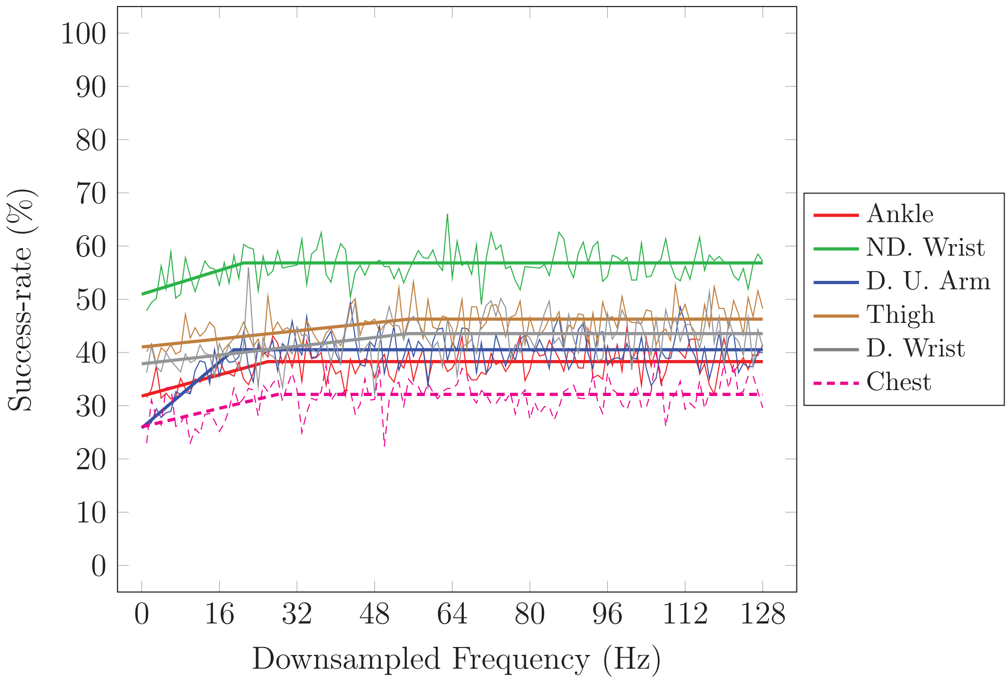
Dusting



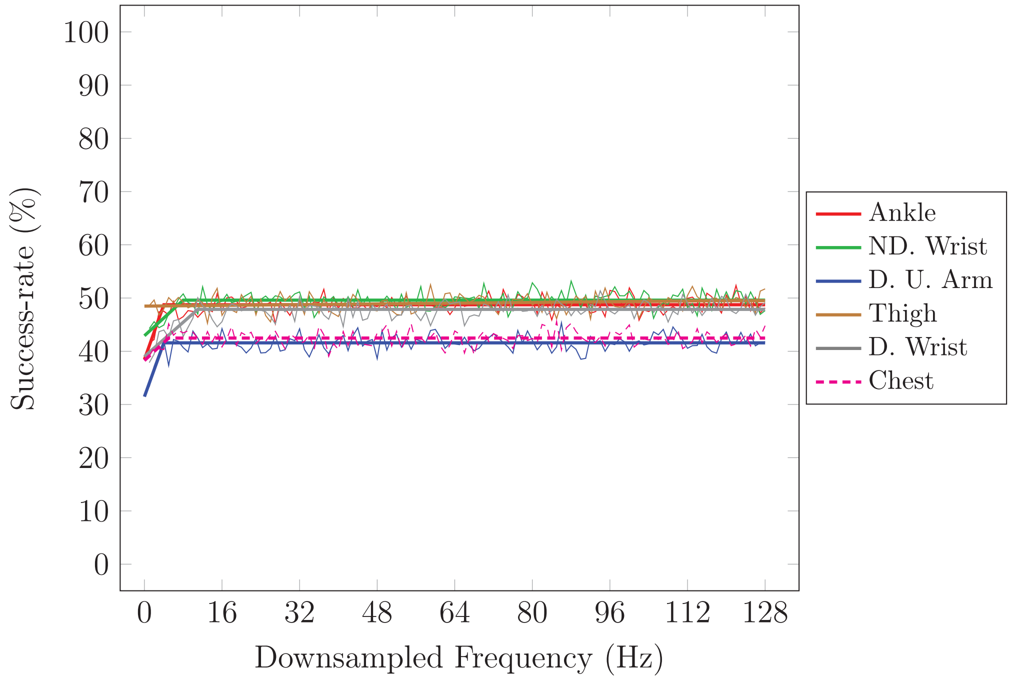
Folding Clothes



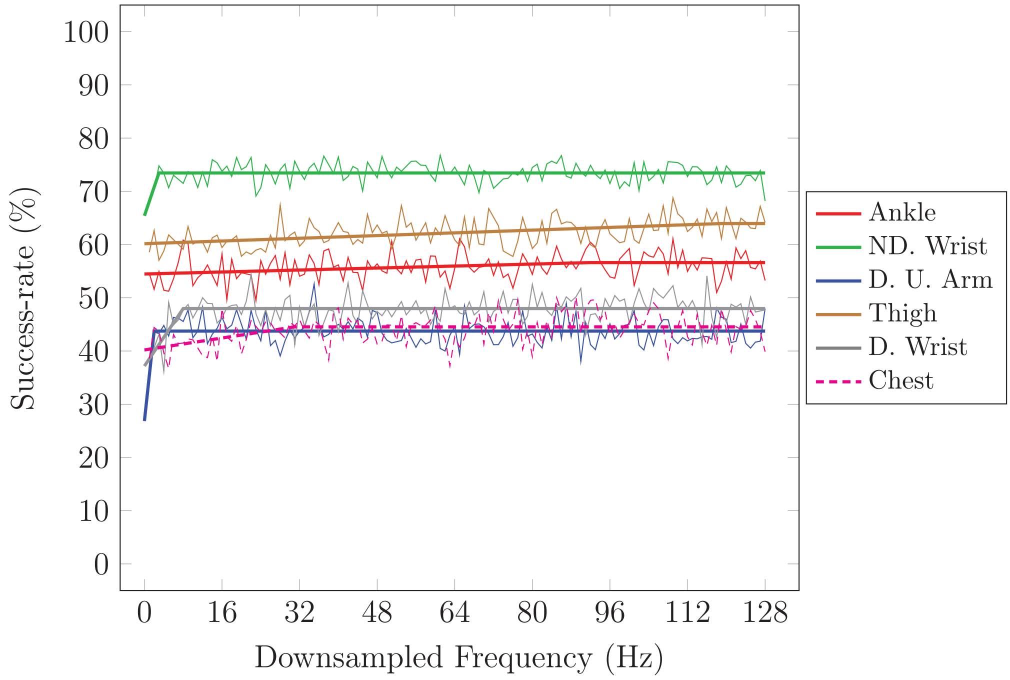
Grating



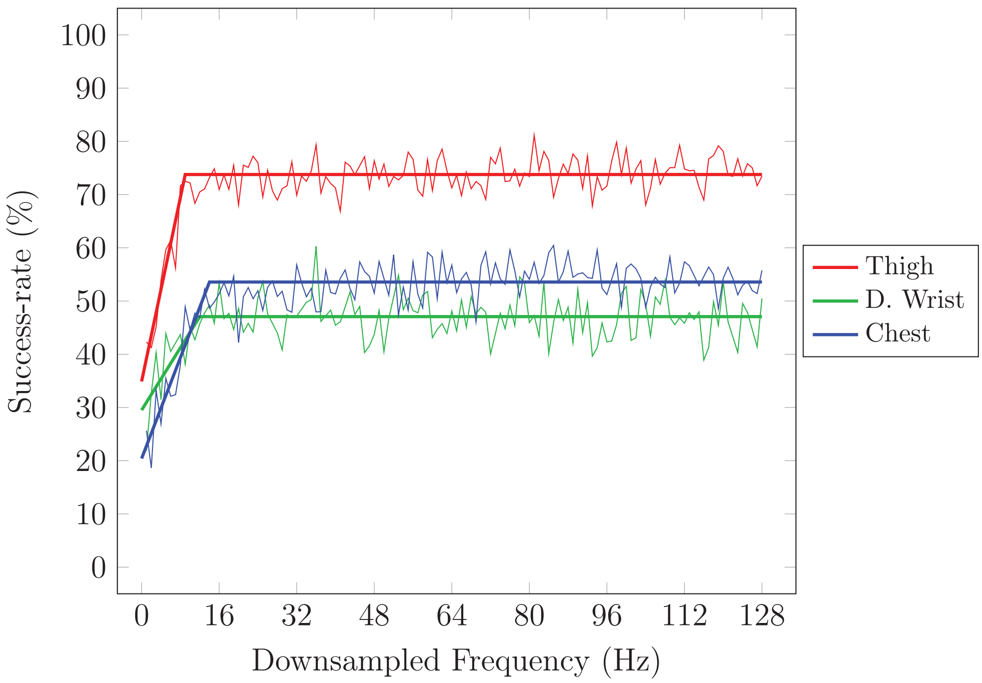
Ironing



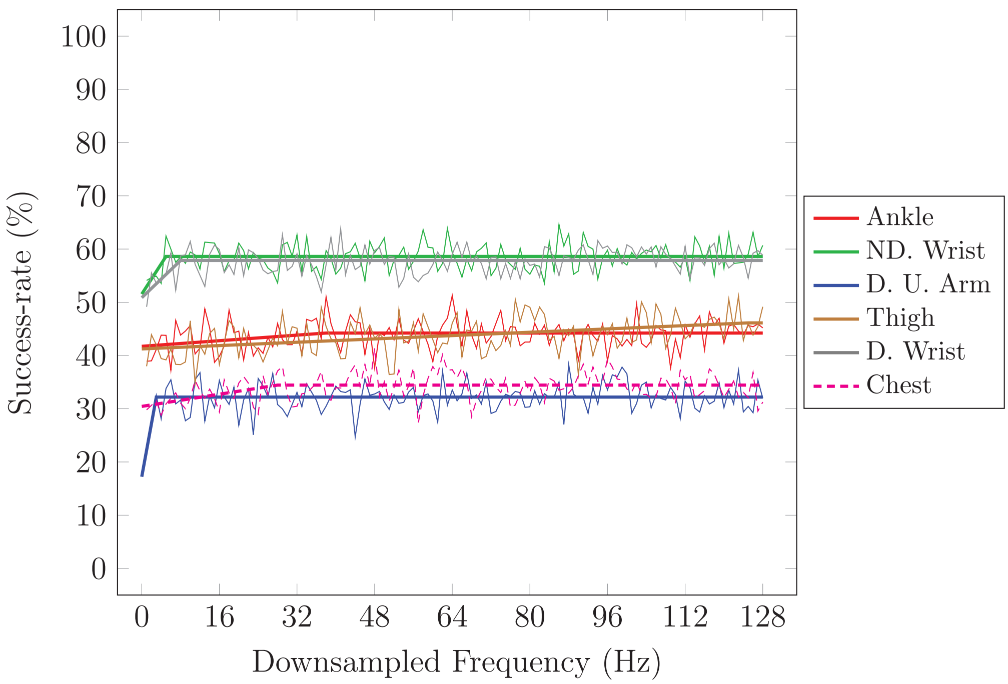
Peeling Vegetables



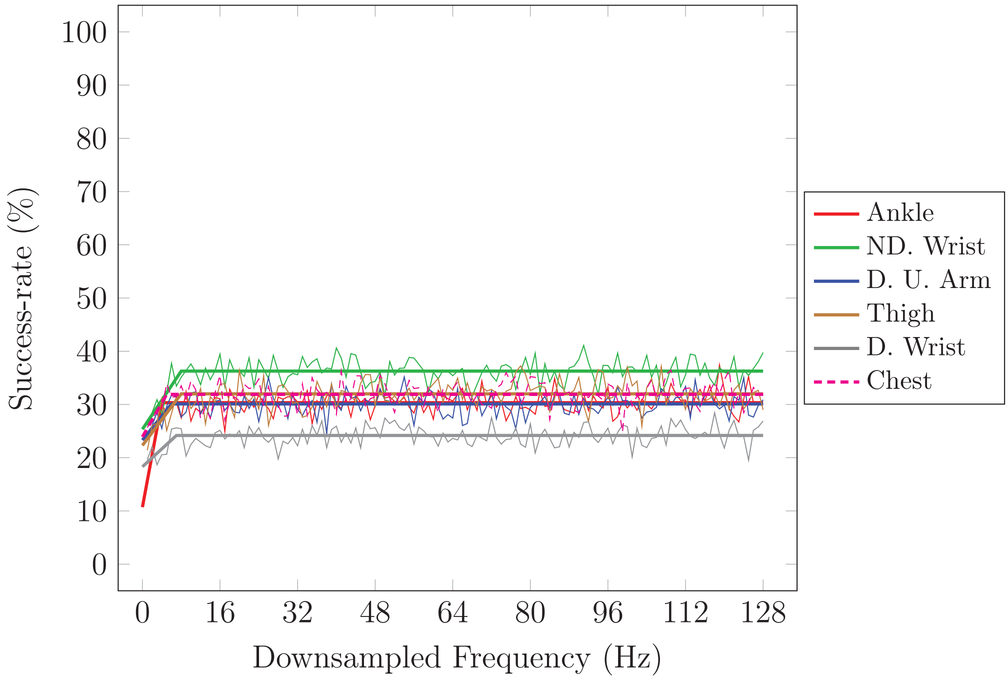
Running



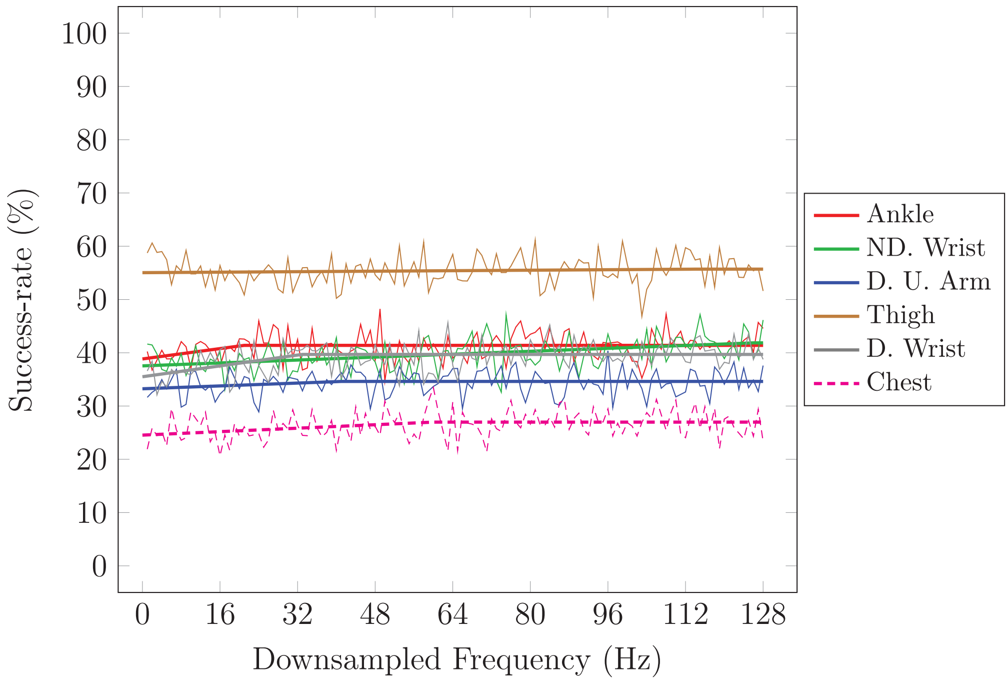
Stiring



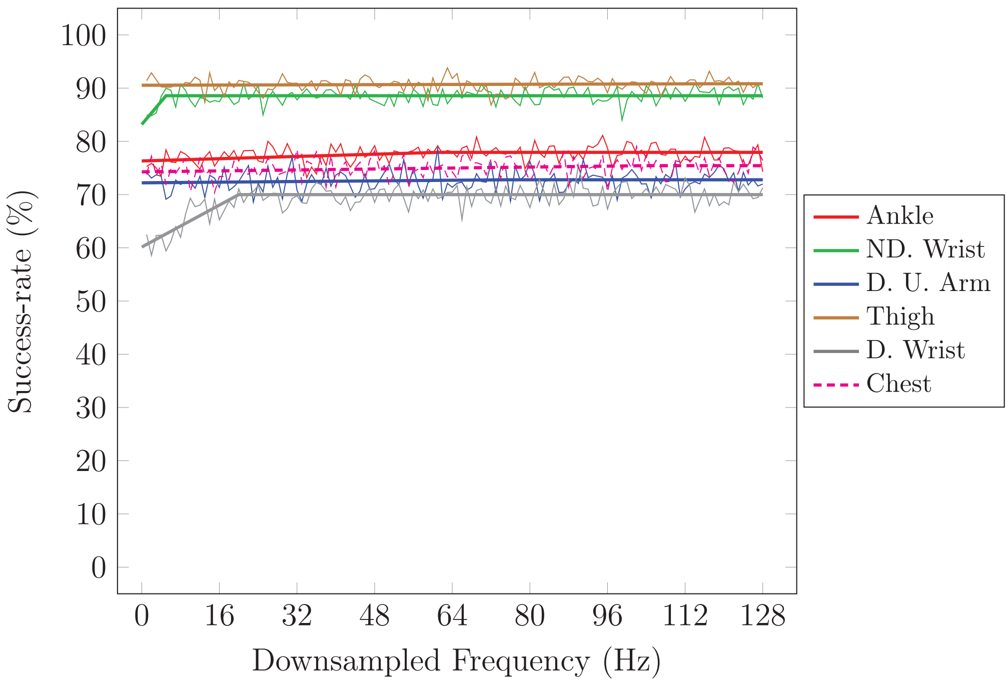
Sweeping



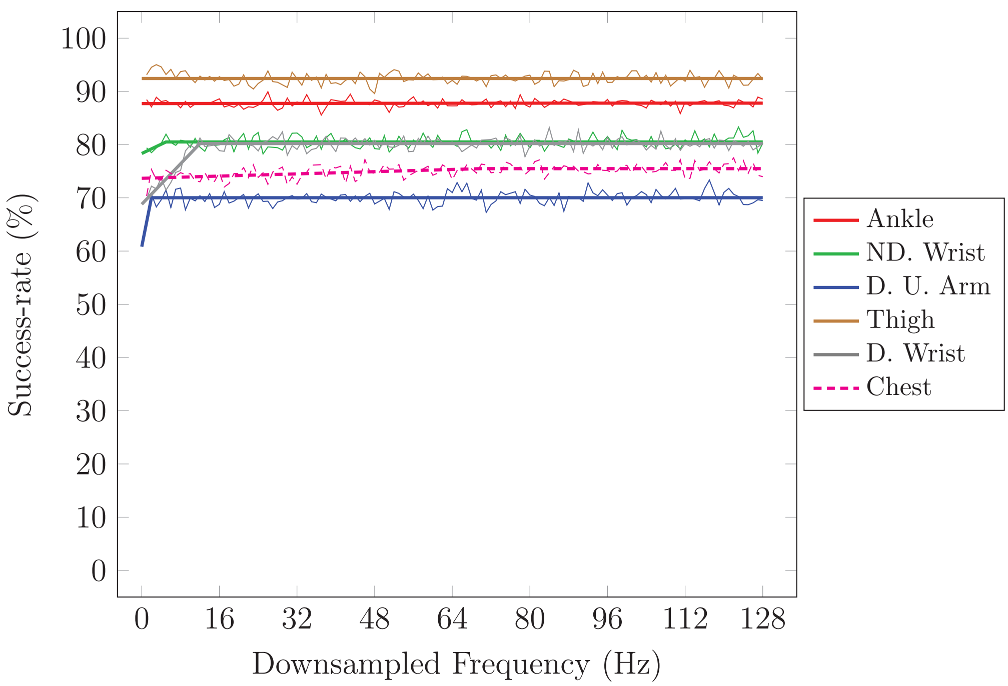
Talking on a Phone



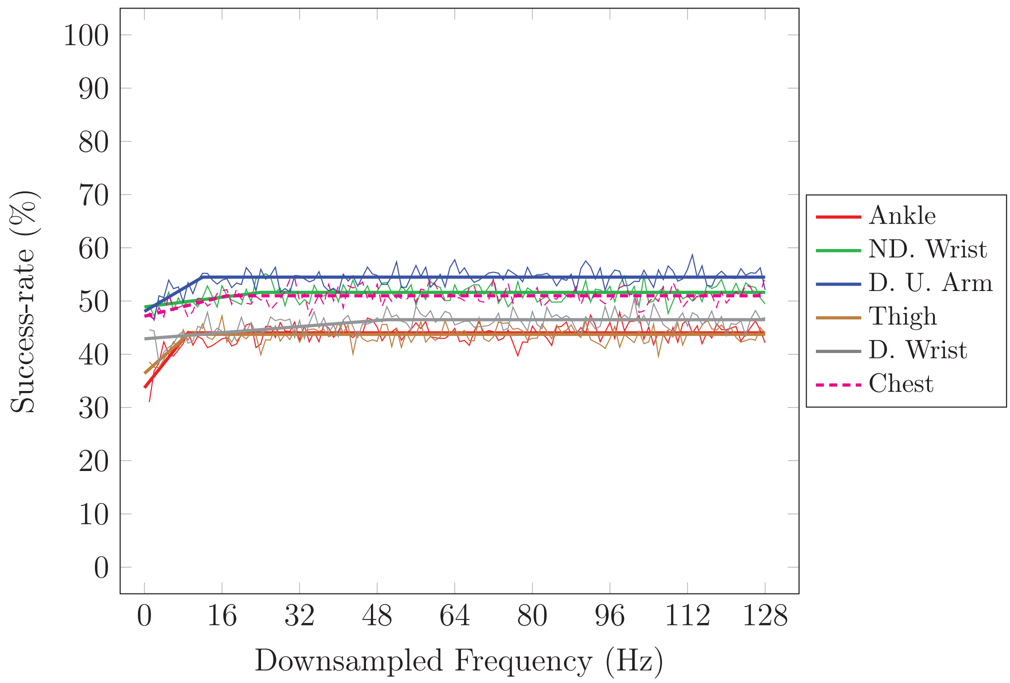
Texting on a Phone



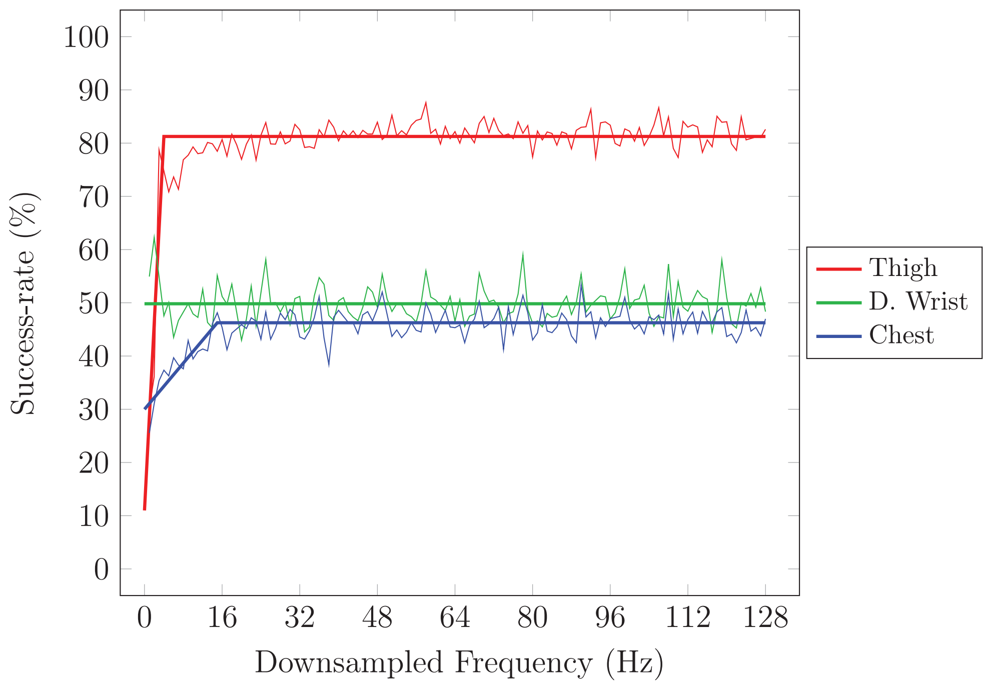
Using a PC



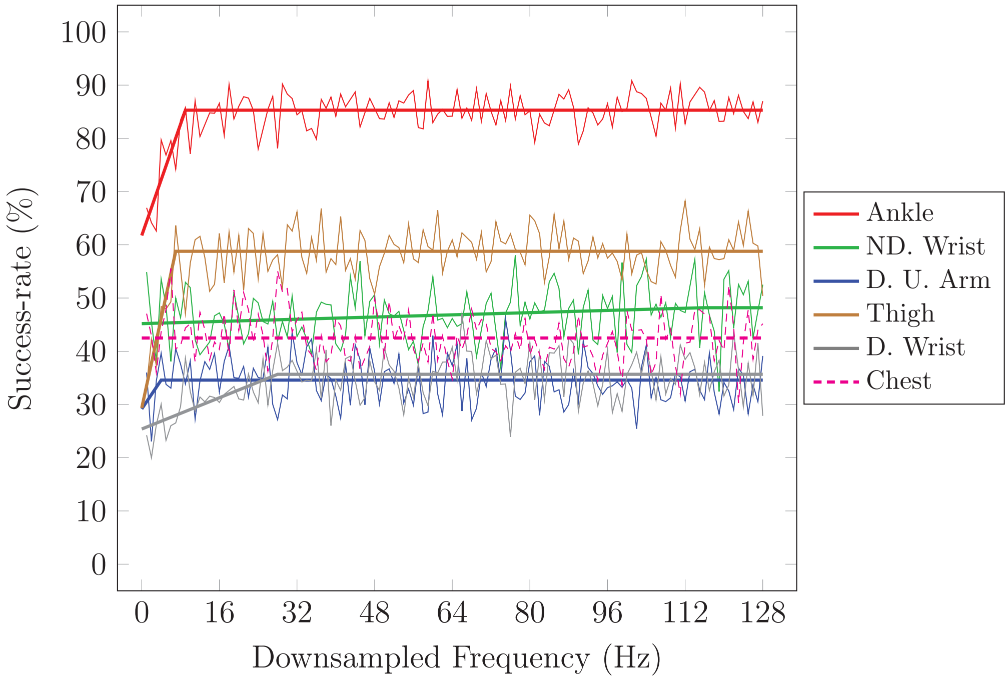
Vacuuming



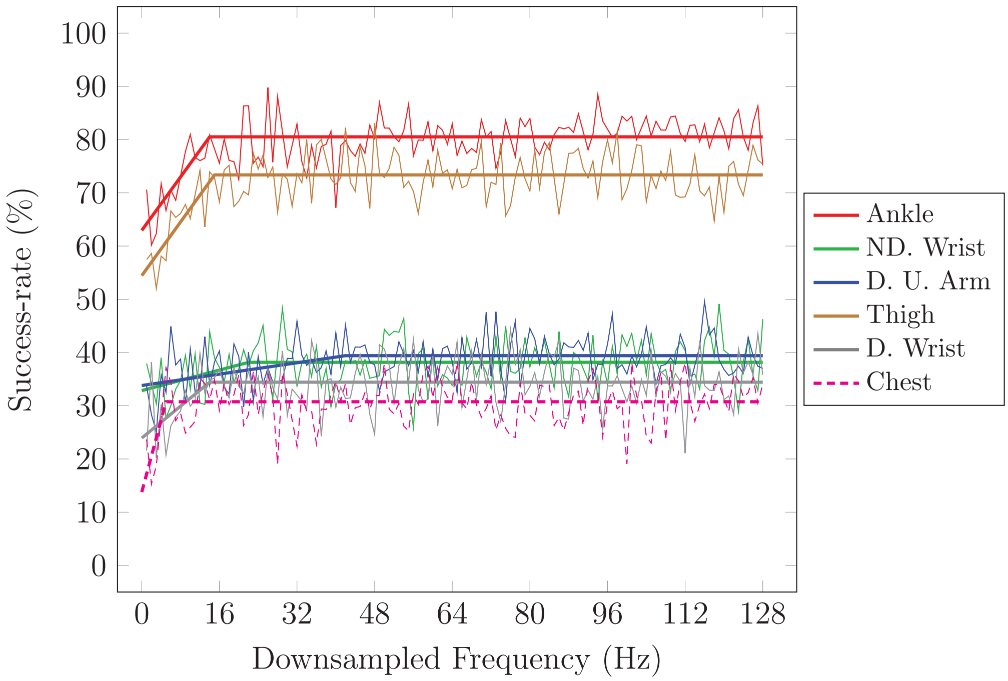
Walking (Flat Ground)



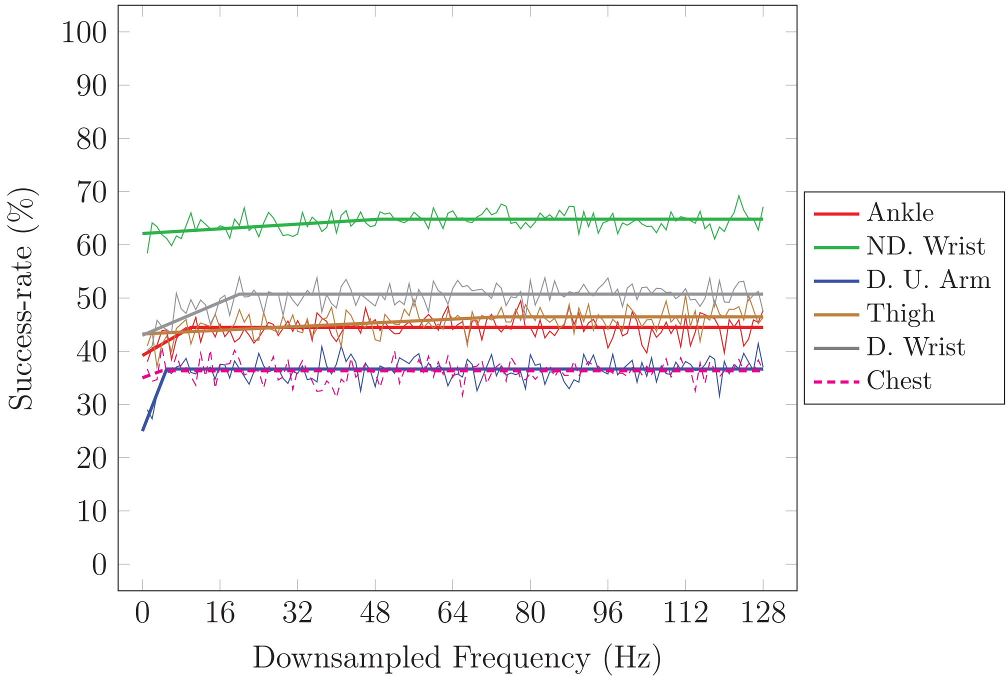
Walking Down Stairs



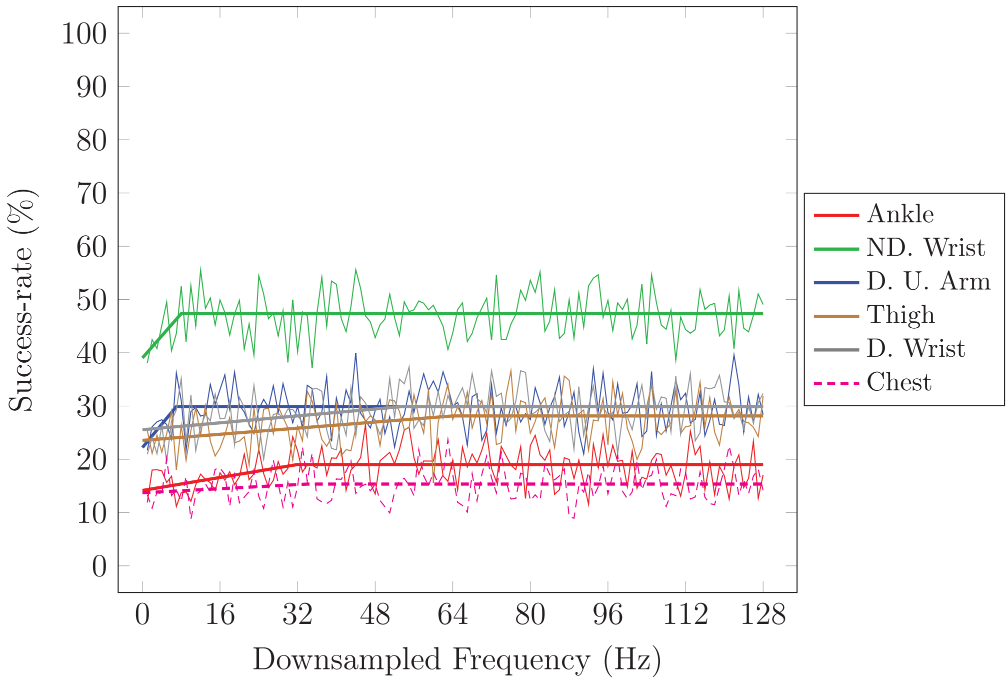
Walking Up Stairs



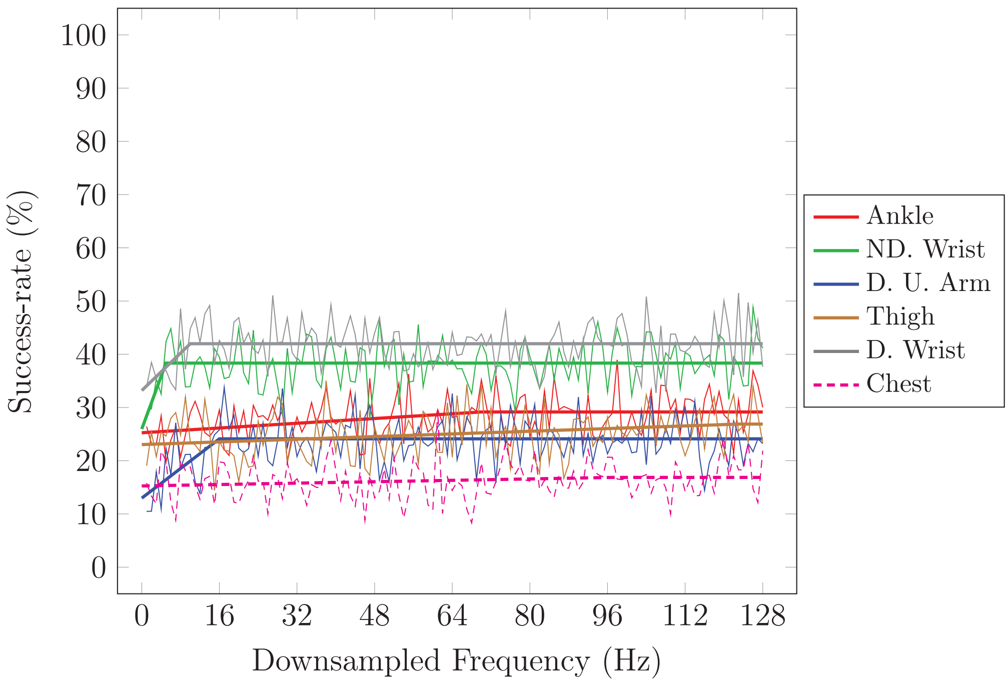
Washing Dishes



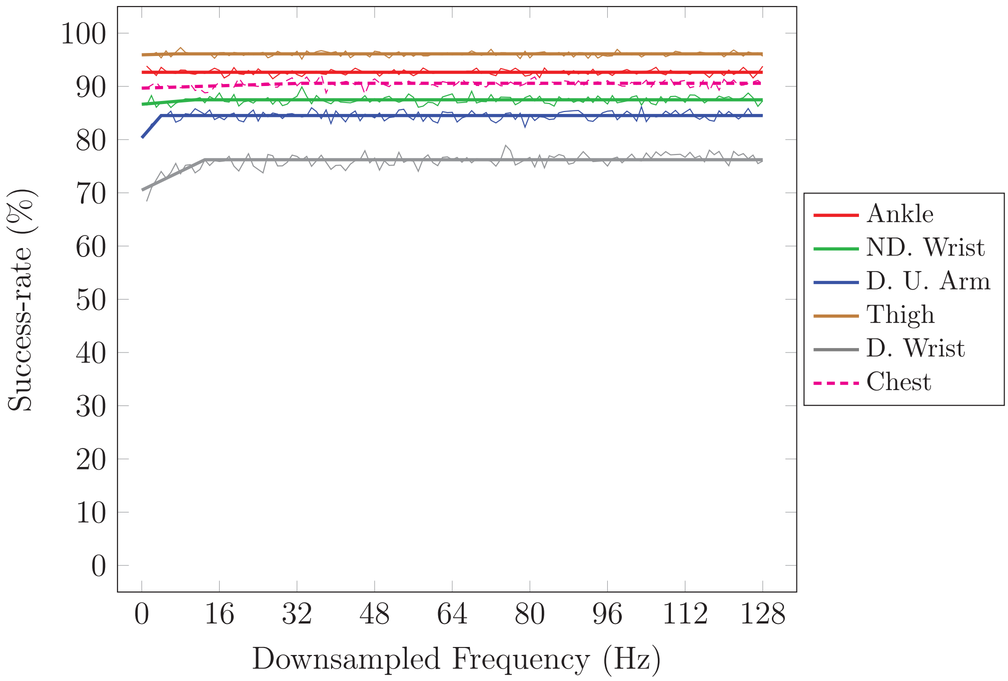
Washing Hands



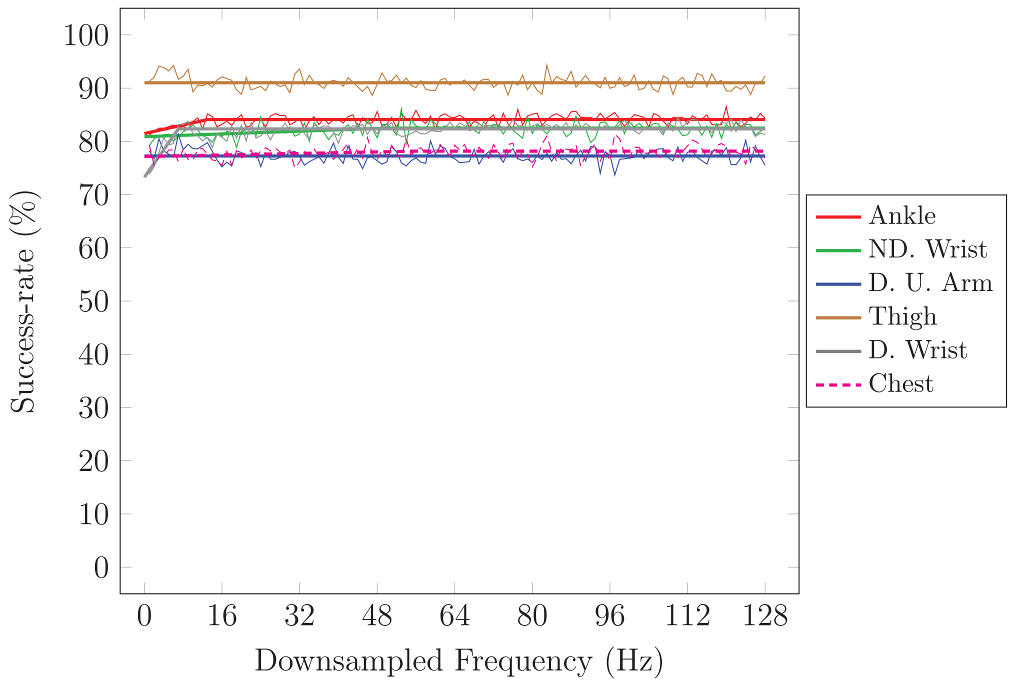
Washing Vegetables



Watching TV



Writing



100
90
80
70
60
50
40
30
20
10
0
Success-rate (%)
0
16
32
48
64
80
96
112
128
Downsampled Frequency (Hz)
Ankle
ND. Wrist
D. U. Arm
Thigh
D. Wrist
Chest

B Minimum Efficient Sampling Frequency on a per location and per activity basis

The MES frequency is defined in section 4.2 as the lowest frequency at which an activity can be sampled from a particular body-location to achieve an activity classification success-rate that is independent of the sampling frequency to a 95% confidence interval. This appendix presents the computed MES frequencies.

First, the MES frequency obtained of each activity, *source* and of the two studied feature-sets is presented.

Then, more details are presented as the MES frequencies of each activity, *source*, monitor and of the two studied feature-sets.

Overall

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	16	16	11	21	< 1	3	11
Dicing	11	10	9	10	< 1	1	7
Dusting	< 1	26	6	5	2	< 1	7
Folding Clothes	5	< 1	6	5	2	8	4
Grating	14	14	52	13	< 1	1	16
Ironing	6	4	6	5	< 1	5	4
Peeling Vegetables	< 1	6	5	4	1	2	3
Running	7	8	8	7	8	8	8
Stiring	3	4	11	6	4	2	5
Sweeping	2	9	6	4	1	5	5
Talking on a Phone	< 1	10	10	12	< 1	< 1	5
Texting on a Phone	12	< 1	29	23	< 1	< 1	11
Using a PC	< 1	< 1	3	4	< 1	< 1	1
Vacuuming	4	27	17	36	2	7	16
Walking (Flat Ground)	4	4	5	10	3	4	5
Walking Down Stairs	3	5	< 1	4	9	7	5
Walking Up Stairs	3	3	< 1	< 1	10	9	4
Washing Dishes	4	< 1	2	3	3	< 1	2
Washing Hands	4	14	2	2	< 1	1	4
Washing Vegetables	34	5	10	9	< 1	1	10
Watching TV	< 1	< 1	13	4	< 1	< 1	3
Writing	8	7	14	11	< 1	< 1	7
Mean	6	8	10	9	2	3	6

Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

Ankle

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	1	3	< 1	10	< 1	< 1	2
Dicing	< 1	2	< 1	18	< 1	< 1	3
Dusting	2	6	< 1	3	4	5	3
Folding Clothes	< 1	10	2	7	2	5	4
Grating	15	5	6	8	< 1	5	7
Ironing	2	5	3	4	1	3	3
Peeling Vegetables	1	< 1	3	10	< 1	< 1	2
Stiring	< 1	24	2	11	< 1	< 1	6
Sweeping	3	4	2	3	3	4	3
Talking on a Phone	< 1	< 1	2	8	< 1	< 1	2
Texting on a Phone	< 1	< 1	4	8	< 1	< 1	2
Using a PC	< 1	12	< 1	14	< 1	< 1	4
Vacuuming	6	21	3	5	3	7	8
Walking Down Stairs	3	5	< 1	4	9	7	5
Walking Up Stairs	3	3	< 1	< 1	10	9	4
Washing Dishes	6	8	3	14	< 1	3	6
Washing Hands	< 1	< 1	1	4	< 1	< 1	< 1
Washing Vegetables	3	< 1	< 1	< 1	< 1	< 1	< 1
Watching TV	< 1	< 1	2	7	< 1	< 1	2
Writing	< 1	27	8	10	< 1	4	8
Mean	2	7	2	7	2	3	4
Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)							

Non-dominant Wrist

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	5	4	16	13	< 1	3	7
Dicing	< 1	3	3	3	< 1	1	2
Dusting	< 1	26	6	5	2	< 1	7
Folding Clothes	5	< 1	6	5	2	8	4
Grating	5	34	23	14	< 1	1	13
Ironing	< 1	7	4	6	< 1	5	4
Peeling Vegetables	< 1	6	5	4	1	2	3
Stiring	< 1	6	6	12	< 1	2	4
Sweeping	3	9	6	4	2	5	5
Talking on a Phone	< 1	< 1	3	8	< 1	< 1	2
Texting on a Phone	6	< 1	12	12	< 1	3	6
Using a PC	8	< 1	8	11	< 1	1	5
Vacuuming	< 1	16	4	5	2	< 1	5
Walking Down Stairs	8	8	3	5	< 1	< 1	4
Walking Up Stairs	6	4	3	3	< 1	< 1	3
Washing Dishes	< 1	< 1	5	7	3	< 1	3
Washing Hands	4	14	2	2	< 1	1	4
Washing Vegetables	34	5	10	9	< 1	2	10
Watching TV	< 1	4	6	5	< 1	< 1	3
Writing	< 1	< 1	9	7	< 1	< 1	3
Mean	4	7	7	7	< 1	2	5

Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

Dominant Upper Arm

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	13	17	9	11	< 1	< 1	8
Dicing	10	16	6	5	1	1	7
Dusting	2	5	5	3	< 1	< 1	3
Folding Clothes	3	7	3	4	< 1	< 1	3
Grating	14	15	11	13	< 1	11	11
Ironing	5	6	5	4	2	3	4
Peeling Vegetables	8	< 1	12	11	1	2	6
Stiring	9	17	11	11	< 1	2	8
Sweeping	2	< 1	4	3	< 1	3	2
Talking on a Phone	< 1	8	10	11	< 1	< 1	5
Texting on a Phone	11	10	6	11	< 1	< 1	6
Using a PC	15	14	14	5	2	2	9
Vacuuming	4	10	4	5	2	7	5
Walking Down Stairs	8	8	5	< 1	< 1	< 1	4
Walking Up Stairs	6	7	< 1	< 1	< 1	< 1	2
Washing Dishes	3	4	1	2	3	4	3
Washing Hands	< 1	< 1	6	< 1	< 1	1	1
Washing Vegetables	< 1	< 1	1	1	< 1	5	1
Watching TV	1	8	8	10	1	3	5
Writing	14	17	13	5	< 1	< 1	8
Mean	6	8	7	6	< 1	2	5

Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

Thigh

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	2	< 1	9	7	< 1	< 1	3
Dicing	< 1	< 1	8	11	< 1	< 1	3
Dusting	3	19	3	4	3	5	6
Folding Clothes	5	6	4	6	< 1	< 1	4
Grating	13	14	9	9	< 1	< 1	8
Ironing	6	4	6	5	< 1	< 1	4
Peeling Vegetables	< 1	< 1	11	7	< 1	2	3
Running	9	11	8	7	8	8	9
Stiring	< 1	6	11	7	< 1	5	5
Sweeping	2	24	6	14	1	5	9
Talking on a Phone	< 1	10	10	12	< 1	< 1	5
Texting on a Phone	12	< 1	29	23	< 1	< 1	11
Using a PC	< 1	< 1	4	4	< 1	< 1	1
Vacuuming	11	27	17	36	4	6	17
Walking (Flat Ground)	4	4	5	10	3	4	5
Walking Down Stairs	6	4	14	4	6	6	7
Walking Up Stairs	5	3	4	2	7	9	5
Washing Dishes	4	5	5	6	2	< 1	4
Washing Hands	< 1	< 1	2	1	1	< 1	< 1
Washing Vegetables	< 1	1	< 1	< 1	< 1	< 1	< 1
Watching TV	< 1	< 1	13	4	< 1	< 1	3
Writing	< 1	< 1	10	5	< 1	< 1	3
Mean	4	6	9	8	2	2	5

Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

Dominant Wrist

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	16	16	11	21	5	7	13
Dicing	11	10	9	10	14	6	10
Dusting	2	6	5	7	< 1	6	4
Folding Clothes	2	< 1	6	4	5	16	6
Grating	16	14	52	6	< 1	< 1	15
Ironing	6	9	4	4	8	8	7
Peeling Vegetables	6	12	8	5	8	5	7
Running	8	7	6	5	5	7	6
Stiring	3	4	11	6	4	3	5
Sweeping	3	63	5	5	1	3	13
Talking on a Phone	5	8	13	18	2	< 1	8
Texting on a Phone	7	4	19	3	10	14	10
Using a PC	5	< 1	3	4	11	10	6
Vacuuming	< 1	38	4	5	8	11	11
Walking (Flat Ground)	3	36	2	< 1	< 1	< 1	7
Walking Down Stairs	9	10	3	4	< 1	6	5
Walking Up Stairs	4	6	< 1	3	29	3	8
Washing Dishes	2	< 1	2	3	6	12	4
Washing Hands	4	3	< 1	< 1	1	< 1	1
Washing Vegetables	< 1	< 1	< 1	< 1	< 1	1	< 1
Watching TV	14	11	16	5	4	9	10
Writing	8	7	14	11	7	6	9
Mean	6	12	9	6	6	6	7

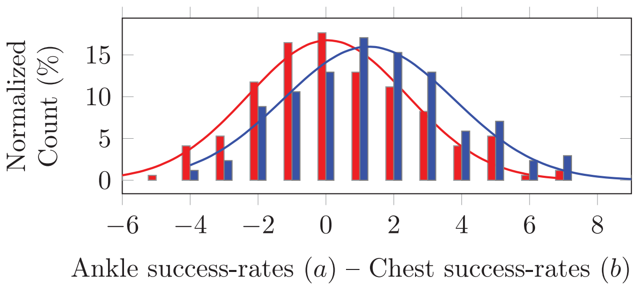
Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

Chest

Activity	Accel.		Rot. Vel.		Orient.		Mean
	B	K	B	K	B	K	
Brushing Teeth	15	18	12	19	2	< 1	11
Dicing	9	25	10	12	< 1	< 1	9
Dusting	4	8	< 1	13	1	< 1	4
Folding Clothes	3	3	< 1	3	< 1	< 1	2
Grating	16	23	22	11	< 1	3	13
Ironing	6	8	< 1	4	< 1	2	3
Peeling Vegetables	< 1	22	11	8	< 1	< 1	7
Running	7	8	6	6	5	12	7
Stiring	< 1	17	16	7	< 1	< 1	7
Sweeping	2	3	< 1	4	1	3	2
Talking on a Phone	1	< 1	8	< 1	< 1	< 1	2
Texting on a Phone	< 1	< 1	13	10	2	< 1	4
Using a PC	16	13	20	15	< 1	< 1	11
Vacuuming	4	7	3	7	< 1	4	4
Walking (Flat Ground)	6	6	3	15	< 1	11	7
Walking Down Stairs	7	7	< 1	2	< 1	< 1	3
Walking Up Stairs	5	6	< 1	< 1	< 1	3	2
Washing Dishes	4	5	3	3	< 1	< 1	3
Washing Hands	< 1	< 1	3	3	< 1	< 1	1
Washing Vegetables	< 1	< 1	< 1	< 1	< 1	< 1	< 1
Watching TV	< 1	< 1	8	10	1	< 1	3
Writing	2	18	16	7	2	< 1	8
Mean	5	9	7	7	< 1	2	5

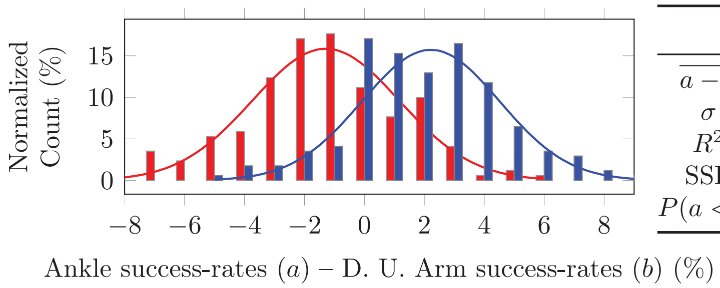
Legend: B:Bao and Intille (2004) K:Kwapisz et al. (2011)

C Comparison of feature sets on a per location basis



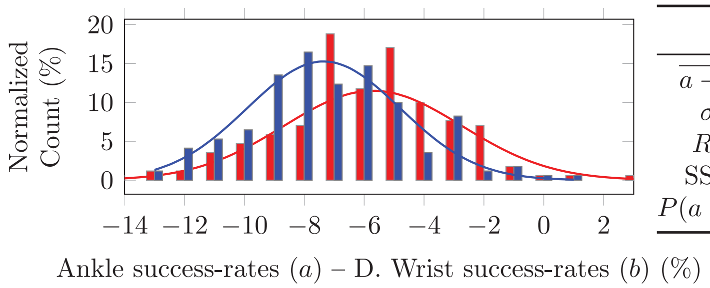
	Bao	***Kwapisz***
$\overline{a-b}$	0.05	1.26
σ	1.8197	1.8862
R^2	0.9378	0.9287
SSE	28.5	27.5
$P(a < b)$	0.4882	0.2523

Figure C.1: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle and classifying feature-vectors extracted from the monitor mounted on the Chest.



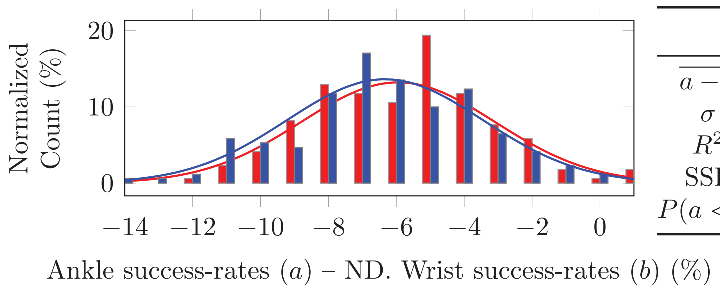
	Bao	*Kwapisz*
$\overline{a-b}$	-1.29	2.22
σ	1.8612	1.8010
R^2	0.8856	0.9224
SSE	53.9	41.5
$P(a<b)$	0.7554	0.1093

Figure C.2: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle and classifying feature-vectors extracted from the monitor mounted on the D. U. Arm.



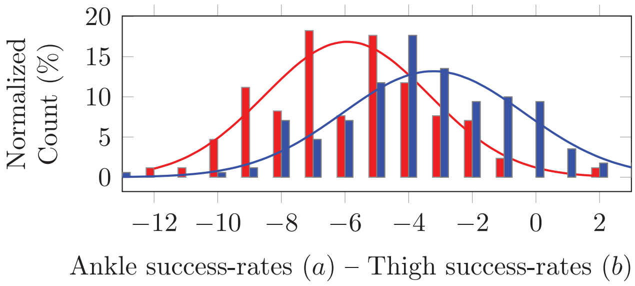
	Bao	*Kwapisz*
$\overline{a-b}$	-5.71	-7.35
σ	2.0508	1.8997
R^2	0.9294	0.8642
SSE	37.3	58.3
$P(a<b)$	0.9973	0.9999

Figure C.3: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle and classifying feature-vectors extracted from the monitor mounted on the D. Wrist.



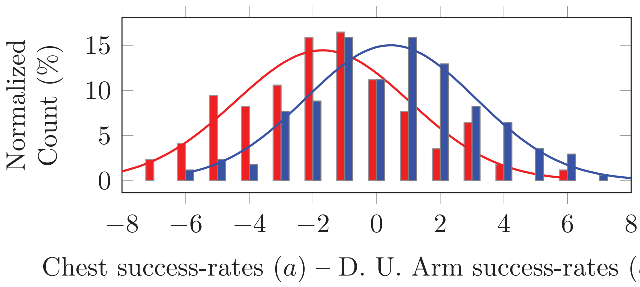
	Bao	*Kwapisz*
$\overline{a-b}$	-5.95	-6.33
σ	2.0106	2.0208
R^2	0.9107	0.8431
SSE	40.8	62.9
$P(a<b)$	0.9985	0.9991

Figure C.4: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle and classifying feature-vectors extracted from the monitor mounted on the ND. Wrist.



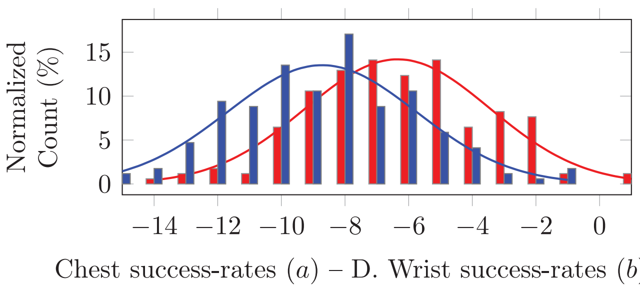
	Bao	*Kwapisz*
$\overline{a-b}$	-5.93	-3.22
σ	1.9110	2.0080
R^2	0.7722	0.8770
SSE	92.0	47.4
$P(a<b)$	0.9990	0.9456

Figure C.5: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Ankle and classifying feature-vectors extracted from the monitor mounted on the Thigh.



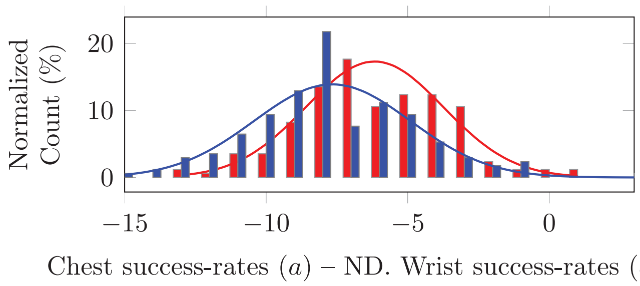
	Bao	*Kwapisz*
$\overline{a-b}$	-1.70	0.44
σ	1.9682	1.9453
R^2	0.8611	0.9199
SSE	47.7	32.6
$P(a<b)$	0.8064	0.4096

Figure C.6: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Chest and classifying feature-vectors extracted from the monitor mounted on the D. U. Arm.



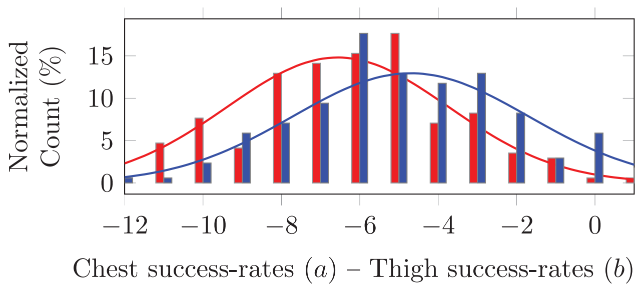
	Bao	*Kwapisz*
$\overline{a-b}$	-6.35	-8.72
σ	2.0240	2.0491
R^2	0.8823	0.8564
SSE	45.0	51.5
$P(a<b)$	0.9991	1.0000

Figure C.7: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Chest and classifying feature-vectors extracted from the monitor mounted on the D. Wrist.



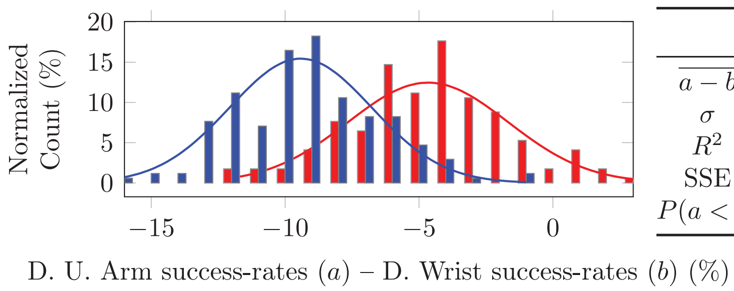
	Bao	***Kwapisz***
$\overline{a-b}$	-6.17	-7.67
σ	1.8542	1.9773
R^2	0.8918	0.9167
SSE	49.9	40.5
$P(a < b)$	0.9996	0.9999

Figure C.8: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Chest and classifying feature-vectors extracted from the monitor mounted on the ND. Wrist.



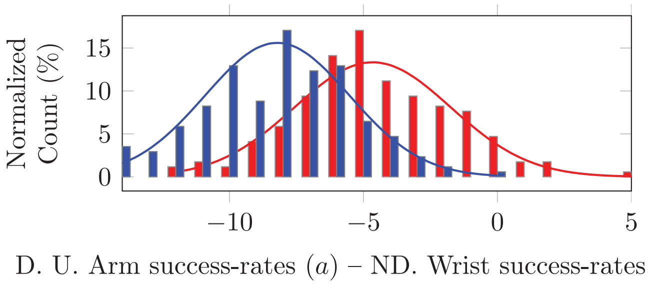
	Bao	***Kwapisz***
$\overline{a-b}$	-6.56	-4.70
σ	1.9976	2.0627
R^2	0.9079	0.9100
SSE	40.0	32.7
$P(a < b)$	0.9995	0.9887

Figure C.9: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the Chest and classifying feature-vectors extracted from the monitor mounted on the Thigh.



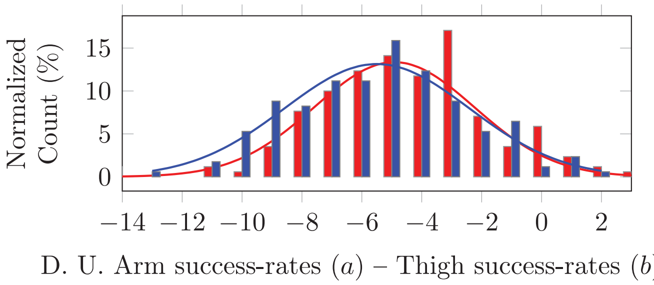
	Bao	***Kwapisz***
$\overline{a-b}$	-4.66	-9.42
σ	2.0557	1.9374
R^2	0.8706	0.8421
SSE	51.1	71.6
$P(a < b)$	0.9883	1.0000

Figure C.10: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. U. Arm and classifying feature-vectors extracted from the monitor mounted on the D. Wrist.



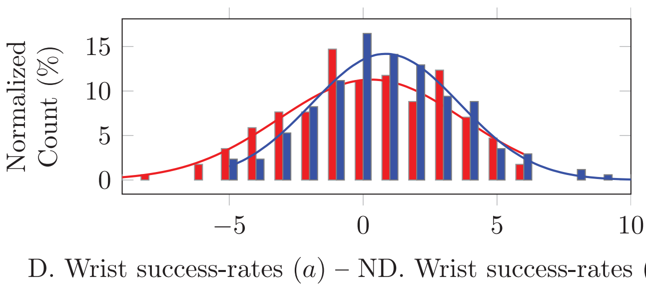
	Bao	**_Kwapisz_**
$\overline{a-b}$	-4.66	-8.22
σ	2.0305	1.9531
R^2	0.9067	0.9044
SSE	35.4	32.2
$P(a<b)$	0.9891	1.0000

Figure C.11: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. U. Arm and classifying feature-vectors extracted from the monitor mounted on the ND. Wrist.



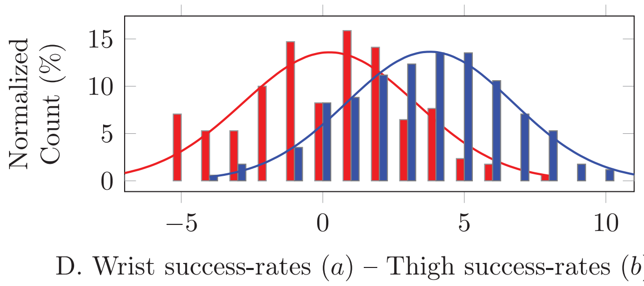
	Bao	**_Kwapisz_**
$\overline{a-b}$	-4.92	-5.50
σ	1.9449	2.0898
R^2	0.9269	0.8941
SSE	31.7	33.8
$P(a<b)$	0.9943	0.9957

Figure C.12: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. U. Arm and classifying feature-vectors extracted from the monitor mounted on the Thigh.



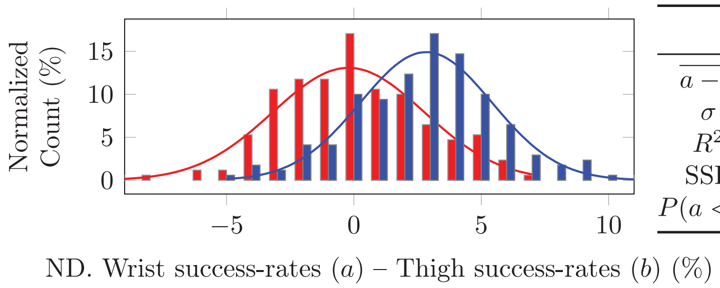
	Bao	**_Kwapisz_**
$\overline{a-b}$	0.30	0.84
σ	2.2134	1.9750
R^2	0.9220	0.9689
SSE	22.3	12.2
$P(a<b)$	0.4464	0.3353

Figure C.13: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. Wrist and classifying feature-vectors extracted from the monitor mounted on the ND. Wrist.



	Bao	***Kwapisz***
$\overline{a-b}$	0.23	3.78
σ	2.0773	2.0383
R^2	0.7398	0.9780
SSE	87.6	7.3
$P(a<b)$	0.4553	0.0317

Figure C.14: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the D. Wrist and classifying feature-vectors extracted from the monitor mounted on the Thigh.



	Bao	***Kwapisz***
$\overline{a-b}$	-0.21	2.84
σ	2.0497	1.9114
R^2	0.9029	0.9308
SSE	38.4	31.7
$P(a<b)$	0.5409	0.0689

Figure C.15: Distribution of the differences between success-rates obtained from classifying feature-vectors extracted from the monitor mounted on the ND. Wrist and classifying feature-vectors extracted from the monitor mounted on the Thigh.

D Activity ranking by likelihood of higher success-rates

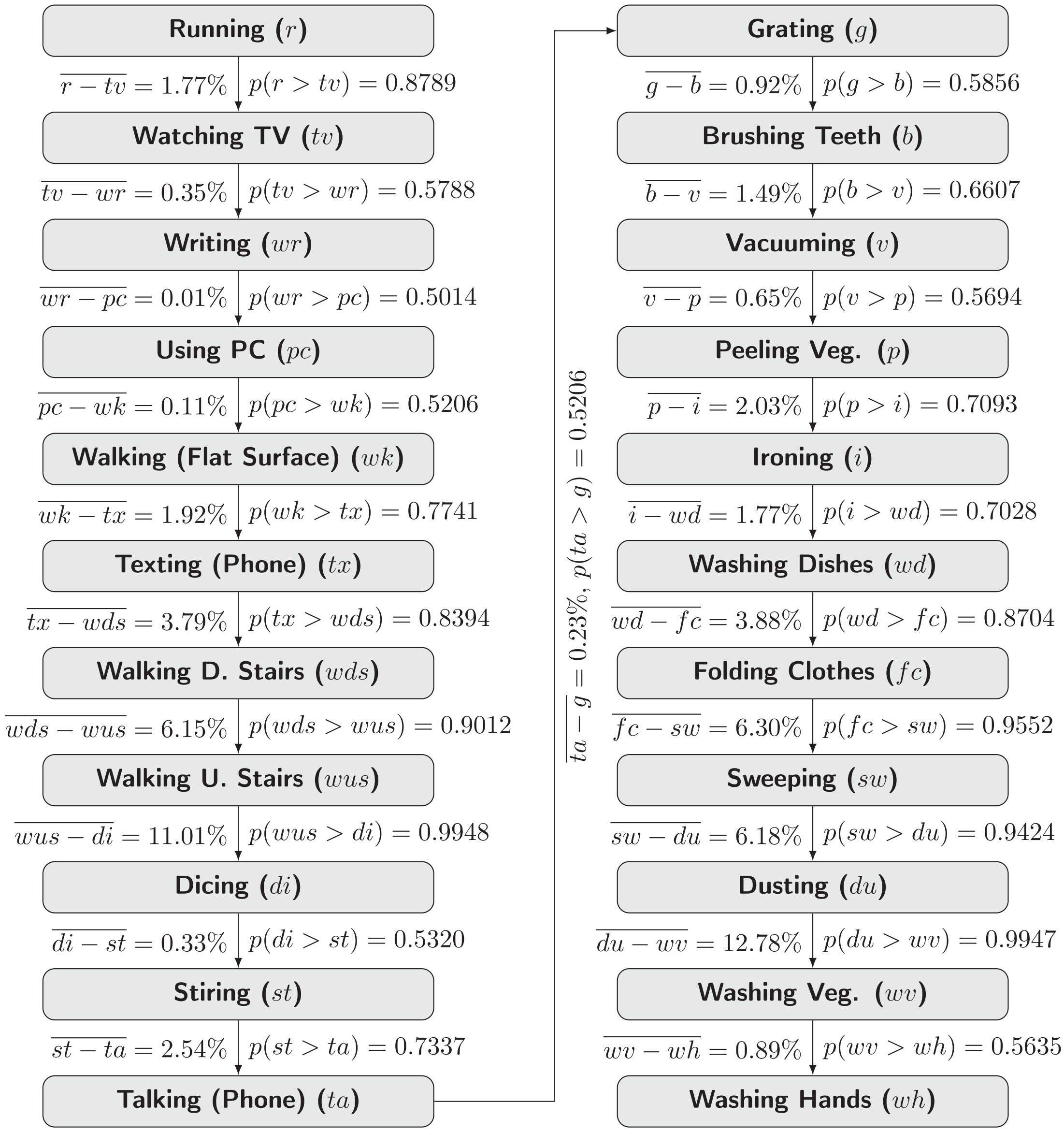


Figure D.1: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. The results are generated from data extracted using Bao and Intille's feature-set using the 3 monitor setup.

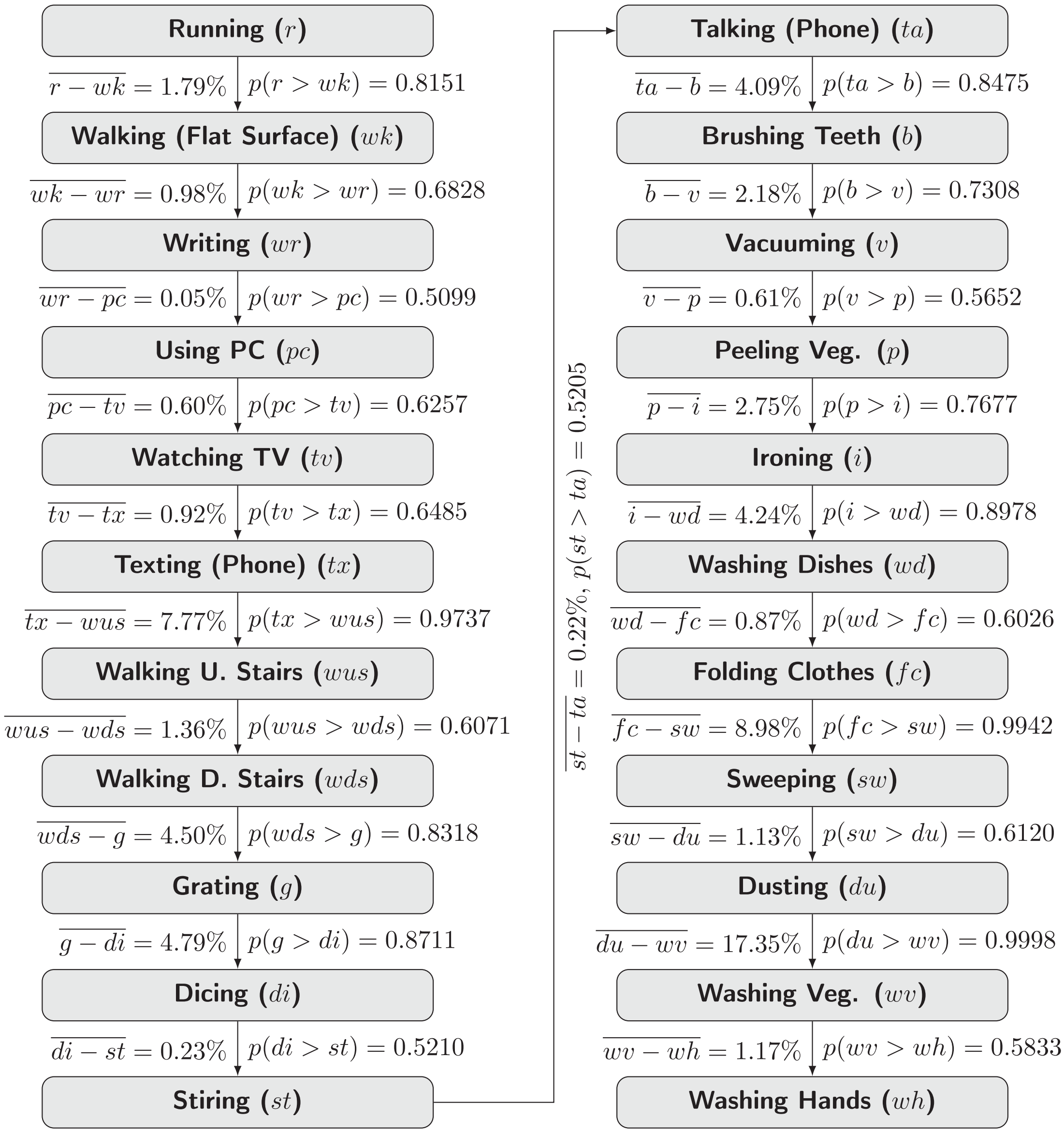


Figure D.2: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. The results are generated from data extracted using Kwapisz et al.'s feature-set using the 3 monitor setup.

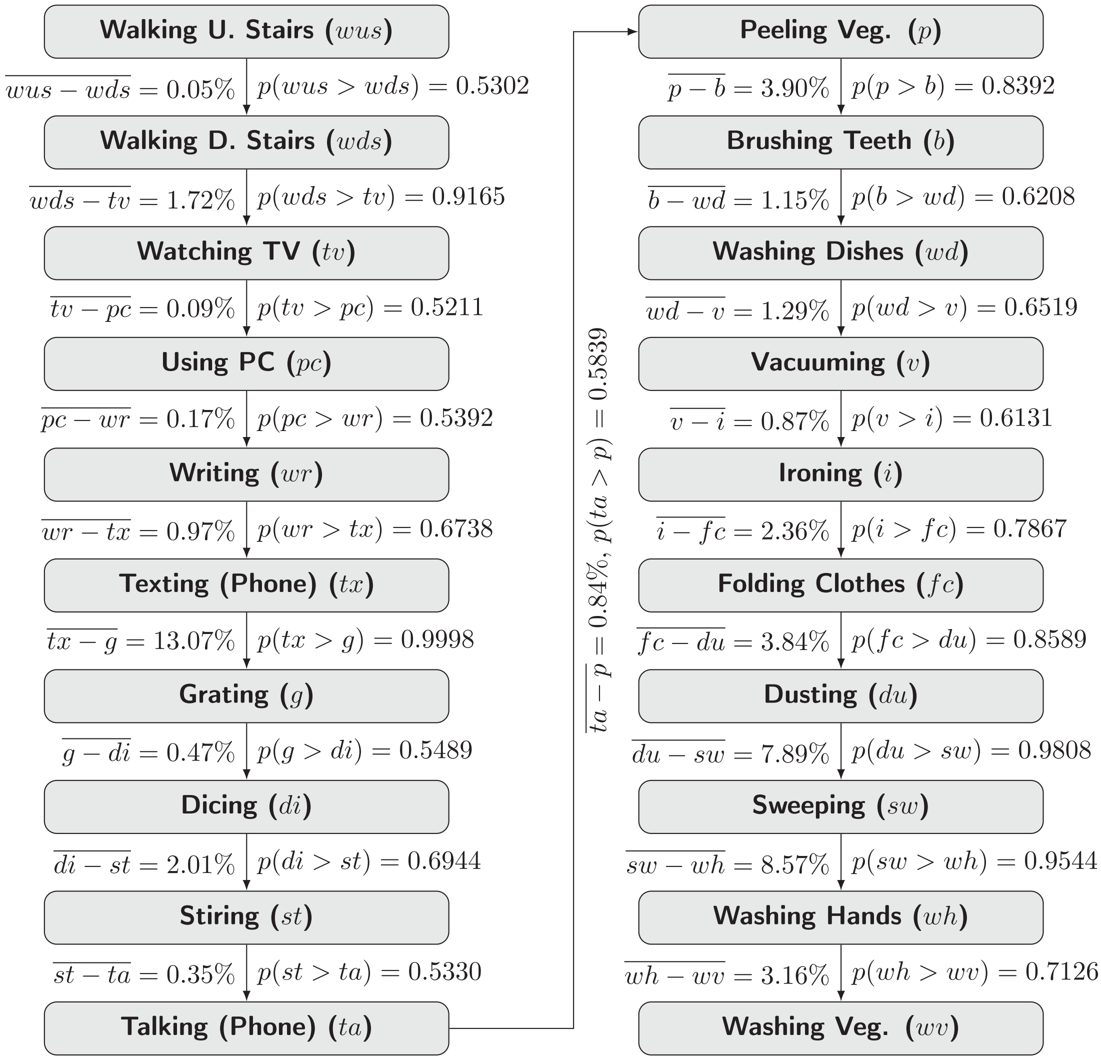


Figure D.3: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. The results are generated from data extracted using Bao and Intille's feature-set using the 6 monitor setup.

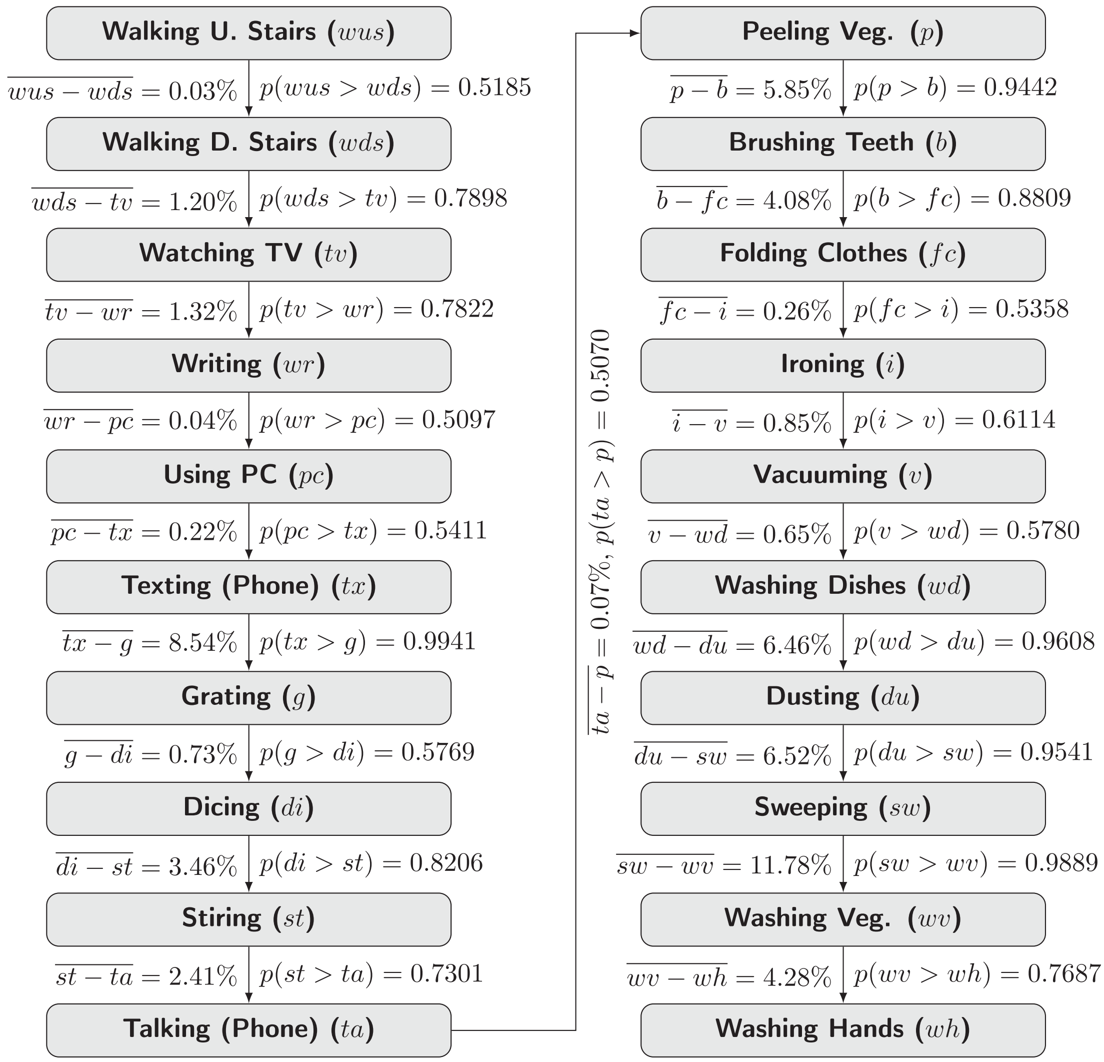
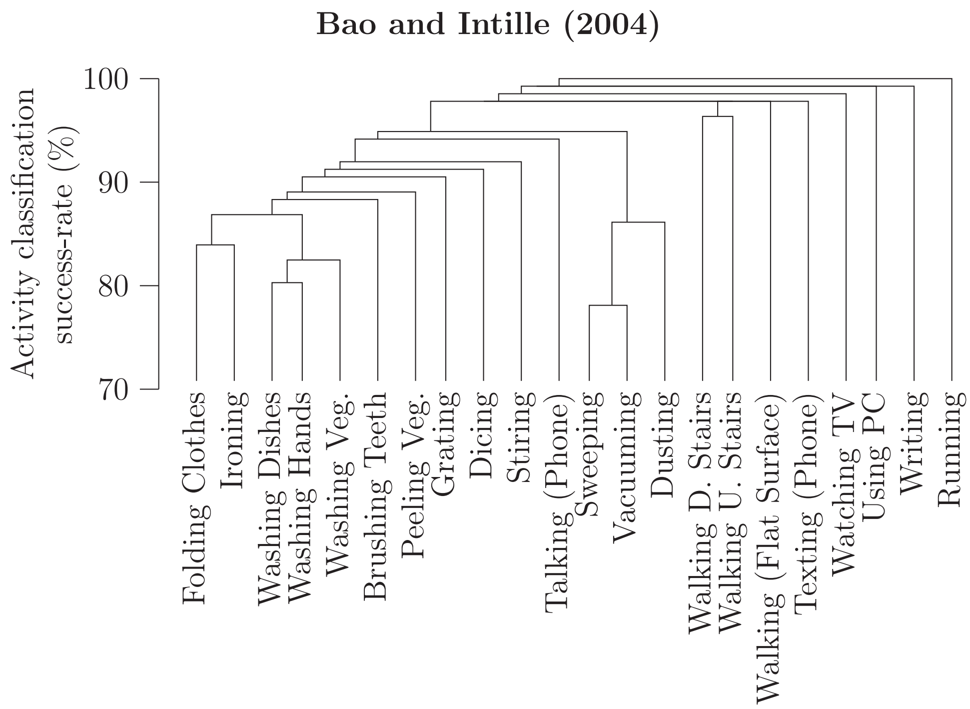


Figure D.4: Illustration showing the ranking of activities based on the likelihood of obtaining higher success-rates for one activity more than other activities. Arrows in activity connections point from activities that are likely to have higher success-rates to those that are likely to have lower success-rates. The results are generated from data extracted using Kwapisz et al.'s feature-set using the 6 monitor setup.

E Activity clustering by mutual error rates

The following are activity dendrograms generated as per the method given in section 6.3. The activity dendrograms present the clustering of activities based on the mutual confusion between activities. Hence, activities with the highest mutual confusion error rates are grouped together first (lowest mergers in the graph). The activities with the least mutual confusion error rates are grouped together last (highest mergers in the graph). The axis on the left presents the activity classification success-rates obtained after combining the data of the activities clustered at that height.



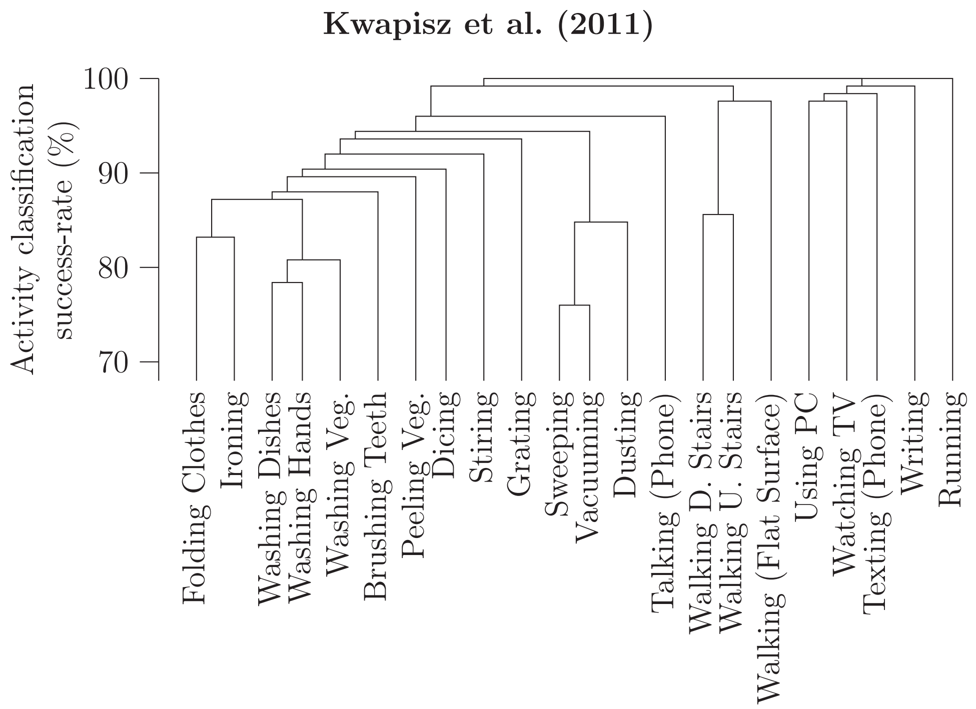
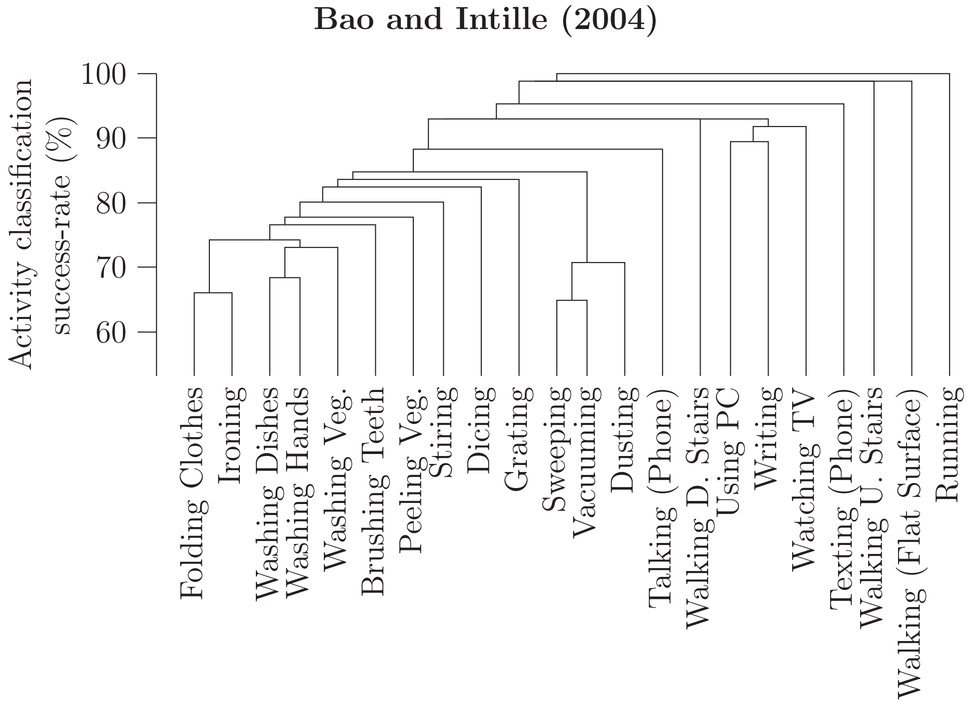


Figure E.1: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Accel-3* (upper) and *Kwapisz-Accel-3* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.



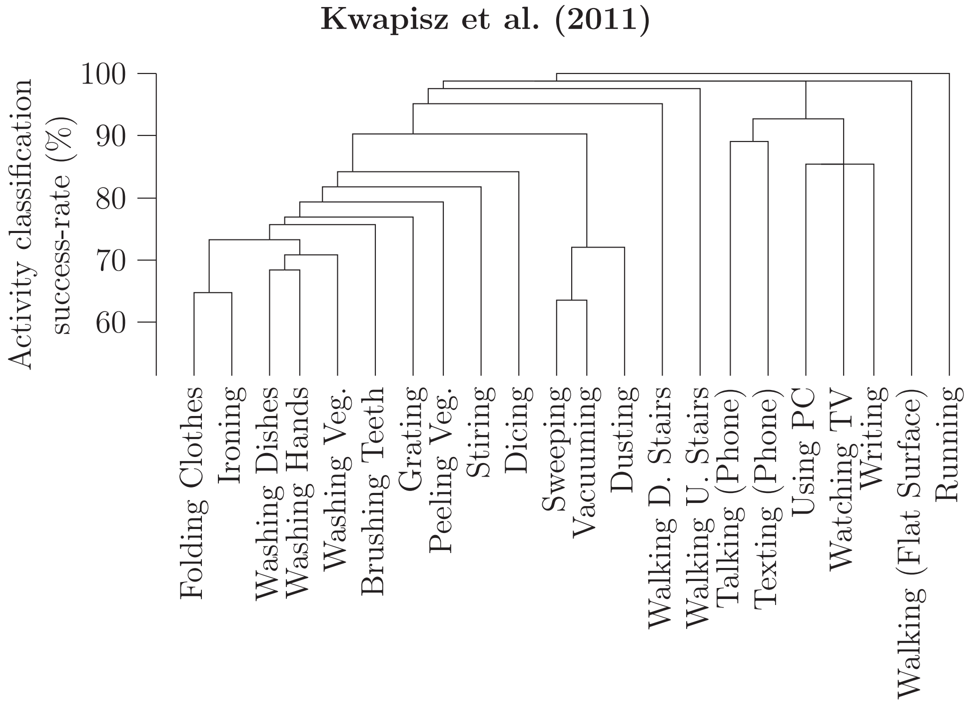
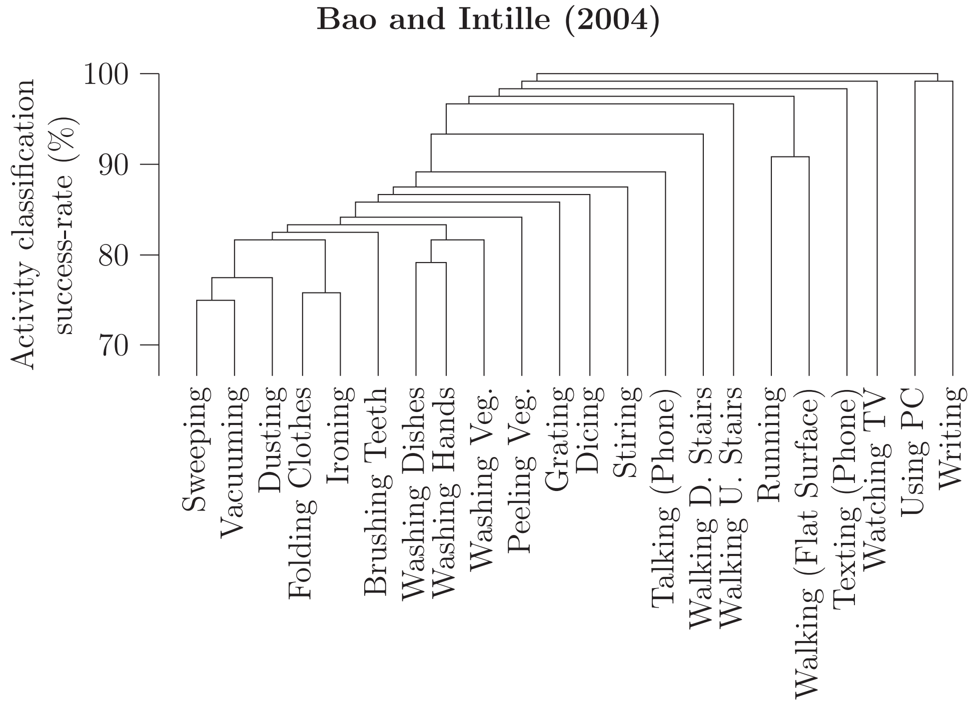


Figure E.2: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Gyro-3* (upper) and *Kwapisz-Gyro-3* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.



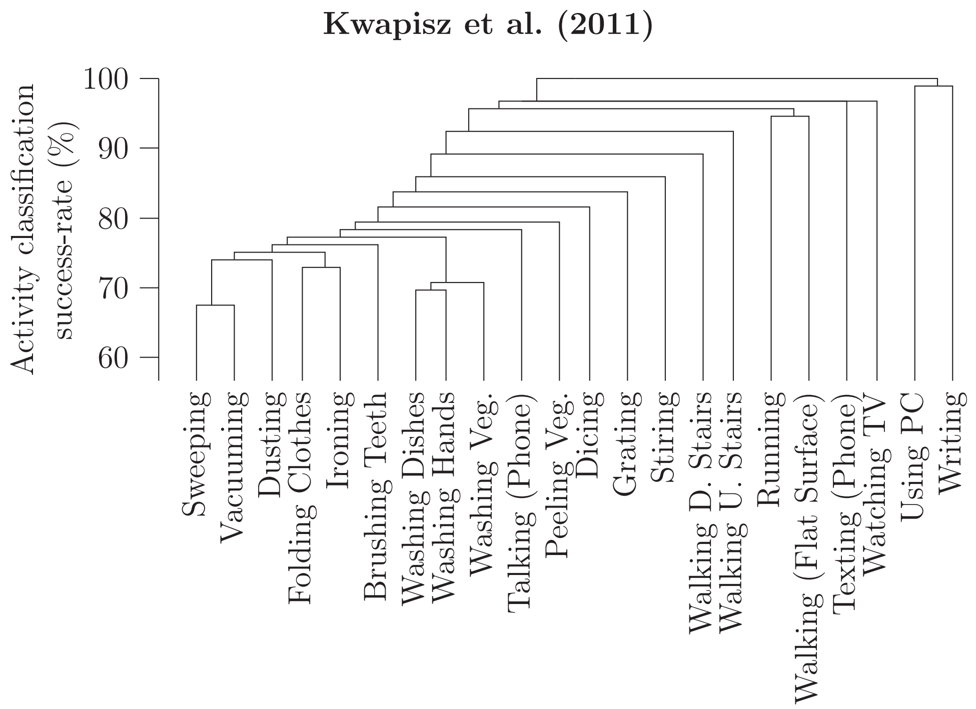
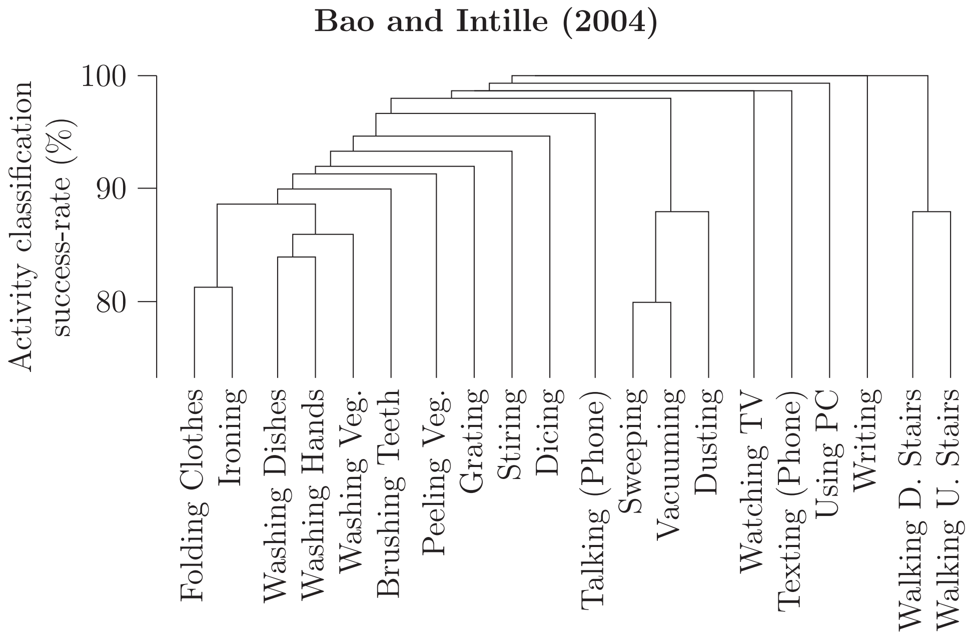


Figure E.3: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Orients-3* (up) and *Kwapisz-Orients-3* (down). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.



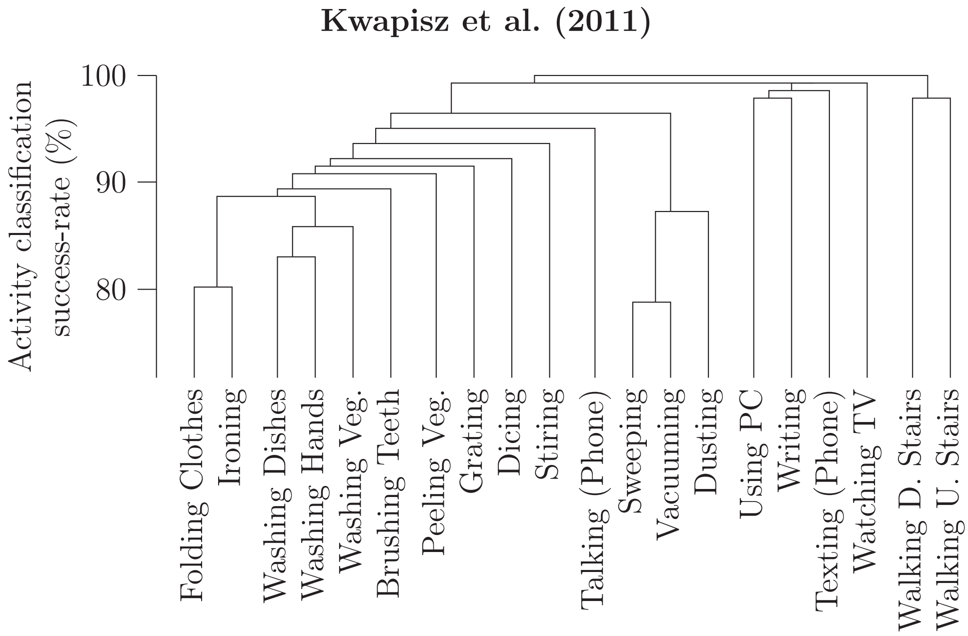


Figure E.4: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Accel-6* (upper) and *Kwapisz-Accel-6* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.

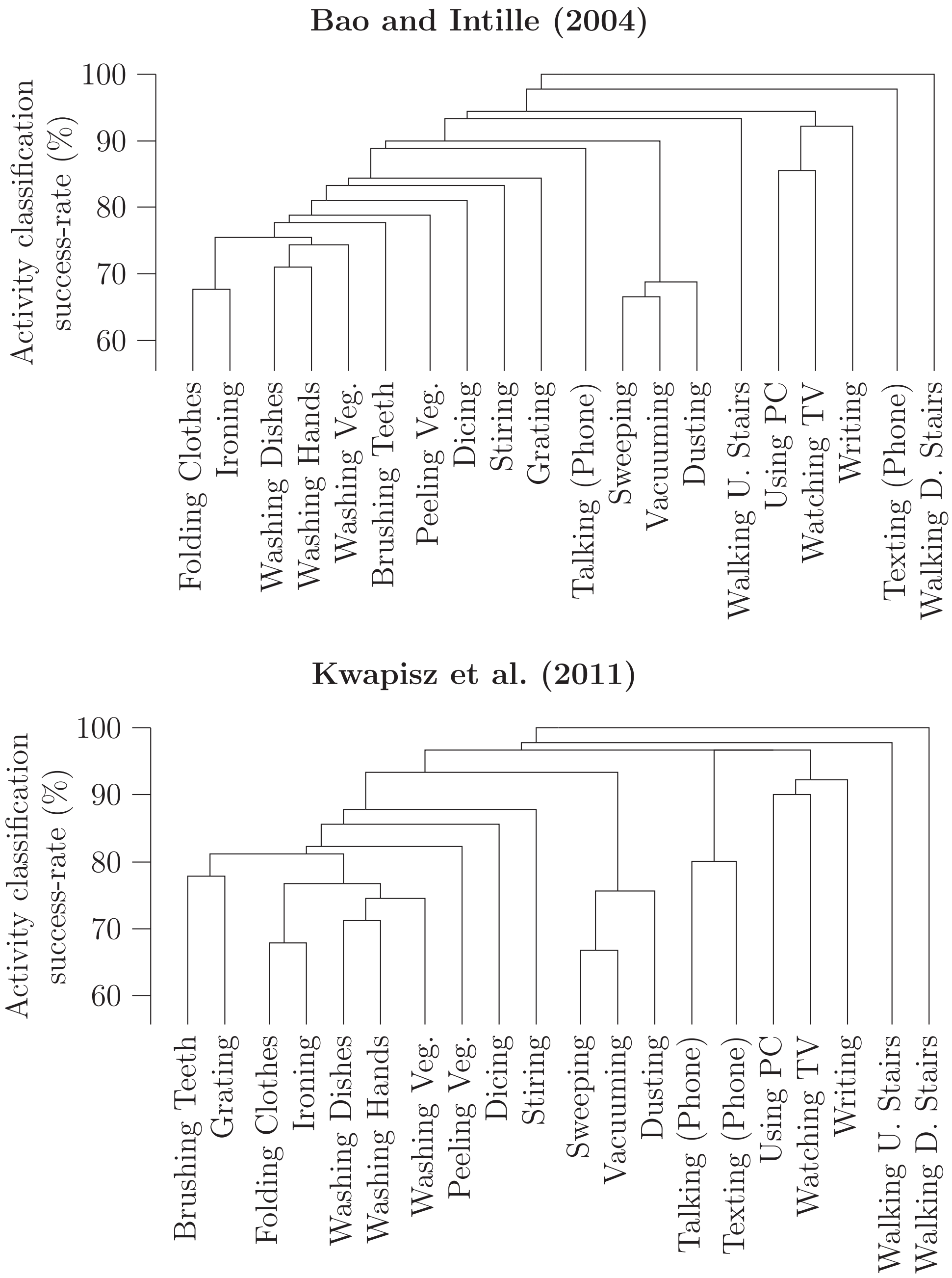
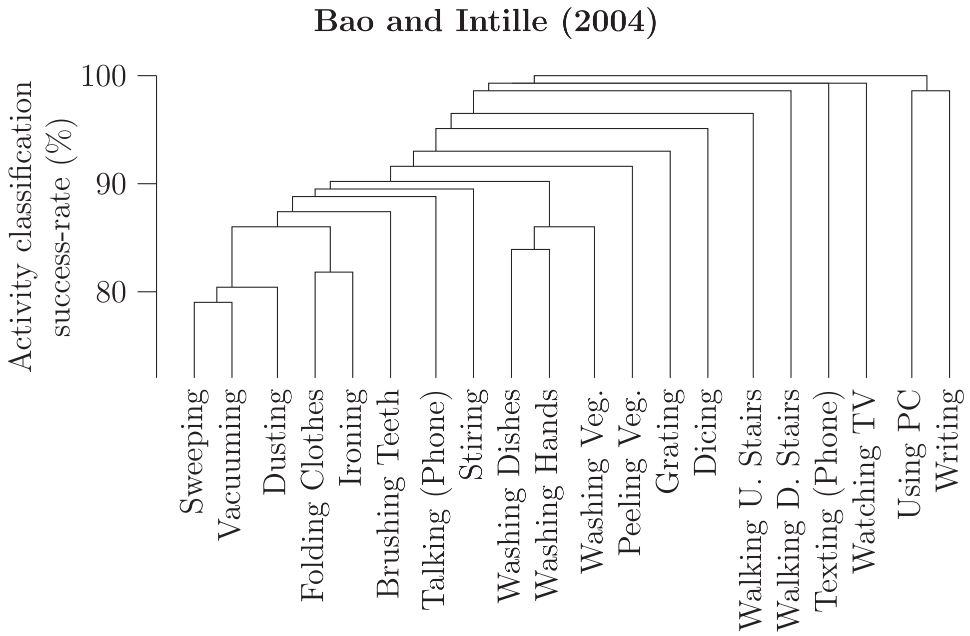


Figure E.5: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Gyro-6* (upper) and *Kwapisz-Gyro-6* (lower). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.



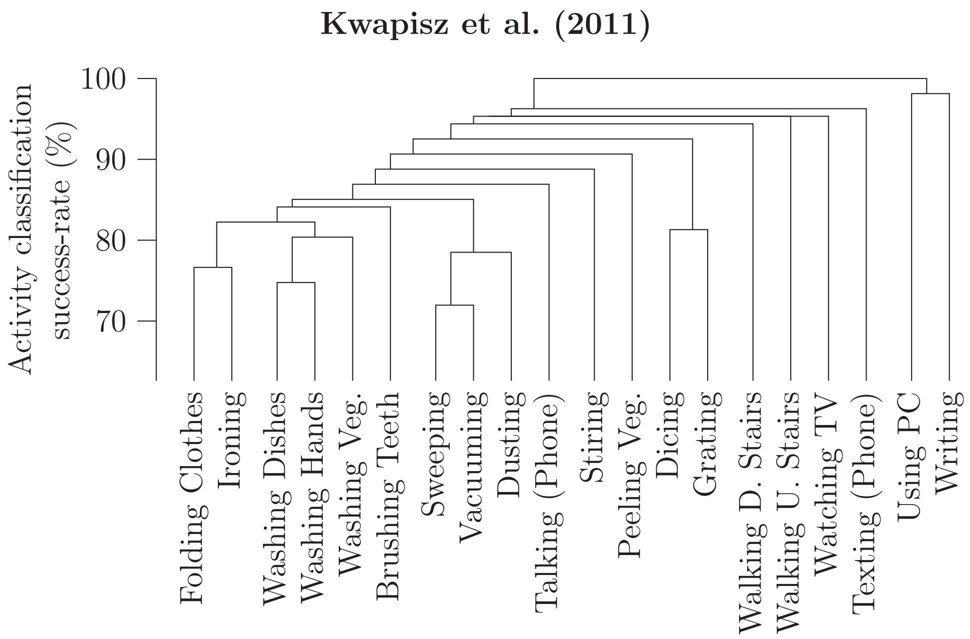


Figure E.6: Activity dendrogram showing clustering of activities by mutual error rates constructed using result sets *Bao-Orients-6* (up) and *Kwapisz-Orients-6* (down). Details of the result sets are given in table 5.4. y-axis shows the success-rates obtained at each activity set merger.

References

Amft, O., & Tröster, G. (2008, feb). Recognition of dietary activity events using on-body sensors. *Artificial intelligence in medicine*, *42*(2), 121–36.

APDM, I. (2012). *Wireless, Wearable, Synchronized Inertial Measurement Units.* Retrieved 19/2/2015, from http://apdm.com/Wearable-Sensors/Opal

Baek, J., Lee, G., Park, W., & Yun, B.-j. (2004). Accelerometer Signal Processing for User Activity Detection. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 610–617). Springer.

Bannach, D., Amft, O., & Lukowicz, P. (2008, apr). Rapid Prototyping of Activity Recognition Applications. *Pervasive Computing, IEEE*, *7*(2), 22–31.

Bao, L. (2003). *Physical Activity Recognition from Acceleration Data under Semi-Naturalistic Conditions* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Bao, L., & Intille, S. (2004). Activity recognition from user-annotated acceleration data. In *2nd International Conference on Pervasive Computing* (pp. 1–17). Springer Berlin Heidelberg.

Beckham, J. (2012). *Fitness trackers use psychology to motivate couch potatoes.* Retrieved 29/5/2014, from http://www.wired.com/2012/04/fitness-tracker-psychology/

Bellotti, V., Price, B., Rasmussen, P., Roberts, M., Schiano, D. J., Walendowski, A., ... Partridge, K. (2008). Activity-based serendipitous recommendations with the Magitti mobile leisure guide. In *Proceedings of the 26th annual CHI conference on Human factors in computing systems - CHI '08* (pp. 1157–1166). New York, New York, USA: ACM Press.

Bieber, G., Voskamp, J., & Urban, B. (2009). Activity Recognition for Everyday Life on Mobile Phones. In *5th International Conference, UAHCI 2009* (pp. 289–296). San Diego, CA: Springer Berlin Heidelberg.

Carós, J. S., Chetelat, O., Celka, P., & Dasen, S. (2005). Very Low Complexity Algorithm for Ambulatory Activity Classification. In *3rd European Medical and Biological Conference EMBEC.* Prague.

Chen, L., Hoey, J., Nugent, C. D., Cook, D. J., Yu, Z., & Member, S. (2012). Sensor-Based Activity Recognition. *IEEE Transactions on Systems, Man, and Cybernetics Society Systems, Man, and Cybernetics Society*, *42*(6), 790–808.

Choudhury, T., Hightower, J., Lamarca, A., Legrand, L., Rahimi, A., Rea, A., ... Wyatt, D. (2008). An Embedded Activity Recognition System. *Pervasive Computing, IEEE*, *7*(2), 32–41.

Consolvo, S., Mcdonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., ... Landay, J. A. (2008). Activity Sensing in the Wild : A Field Trial of UbiFit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1797–1806). ACM digital library.

Cui, Y., Chipchase, J., Ichikawa, F., & Tokyo, N. D. (2007). A Cross Culture Study on Phone Carrying and Physical Personalization. *Usability and Internationalization. HCI and Culture*, *4559*, 483–492.

Dernbach, S., Das, B., Krishnan, N. C., Thomas, B. L., & Cook, D. J. (2012, jun). Simple and Complex Activity Recognition through Smart Phones. In *2012 8th International Conference on Intelligent Environments* (pp. 214–221). Guanajuato, Mexico: IEEE.

Dey, A. (1999). The conference assistant: combining context-awareness with wearable computing. In *3rd International Symposium on Wearable Computers* (pp.

21–28). San Francisco: IEEE.

Ermes, M., & Juha, P. (2008). Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions. *IEEE Transactions on Information Technology in Biomedicine*, *12*(1), 20–26.

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. a. M. P. (2010, mar). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, *14*(7), 645–662.

Ghasemzadeh, H., Barnes, J., Guenterberg, E., & Jafari, R. (2008, sep). A phonological expression for physical movement monitoring in body sensor networks. *2008 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, 58–68.

Gu, T., Wu, Z., Wang, L., Tao, X., & Lu, J. (2009). Mining Emerging Patterns for recognizing activities of multiple users in pervasive computing. In *Proceedings of the 6th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (pp. 1–10). Toronto, Ontario: IEEE.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, *11*(1).

Hanai, Y., Nishimura, J., & Kuroda, T. (2009). Haar-like filtering for human activity recognition using 3D accelerometer. *Machine Learning*, *512*(1), 675–678.

He, Z., Liu, Z., Jin, L., Zhen, L.-x., & Huang, J.-c. (2008). Weightlessness Feature âĂŤ A Novel Feature for Single Tri-axial Accelerometer based Activity Recognition. In *19th International Conference on Pattern Recognition* (pp. 1–4). IEEE.

Henpraserttae, A., Thiemjarus, S., & Marukatat, S. (2011). Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation and Location. In *2011 International Conference on Body Sensor Networks* (pp. 41–46). IEEE.

Huynh, T., & Schiele, B. (2005). Analyzing features for activity recognition. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies* (pp. 159 – 163). New York, New York, USA.

Jeon, A. Y., Kim, J. H., Kim, I. C., Ye, S. Y., Ro, J. H., Yoon, S. H., . . . Jeon, G. R. (2007). Implementation of the Personal Emergency Response System using a 3-axial Accelerometer. In *6th International Special Topic Conference on Information Technology Applications in Biomedicine* (pp. 223–226). IEEE.

Keally, M., Xing, G., & Pyles, A. (2011). PBN : Towards Practical Activity Recognition Using Smartphone-Based Body Sensor Networks âĹŮ. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems.* New York, New York, USA: ACM digital library.

Kern, N., Schiele, B., & Schmidt, A. (2003). Multi-Sensor Activity Context Detection for Wearable Computing. *Ambient Intelligence*, 220–232.

Khan, A. M., Lee, Y. K., & Kim, T. (2008). Accelerometer Signal-based Human Activity Recognition Using Augmented Autoregressive Model Coefficients and Artificial Neural Nets. In *30th Annual International Conference of Engineering in Medicine and Biology Society* (pp. 5172–5175). Vancouver, BC: IEEE.

Kim, E., Helal, S., & Cook, D. J. (2010). Human activity recognition and pattern discovery. *Pervasive Computing, IEEE*, *9*(1), 48–53.

King, R. C., Mcilwraith, D. G., Lo, B., Pansiot, J., Mcgregor, A. H., & Yang, G.-z. (2009). Body Sensor Networks for Monitoring Rowing Technique. In *6th International Workshop on Wearable and Implantable Body Sensor Networks* (pp. 251–255). IEEE.

Korel, B. T., & Koo, S. G. M. (2007). Addressing Context Awareness Techniques in Body Sensor Networks. In *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07).*

Krishnan, N., & Panchanathan, S. (2008). Analysis of low resolution accelerometer data for continuous human activity recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008* (pp. 3337–3340).

Kunze, K., & Lukowicz, P. (2008). Dealing With Sensor Displacement In Motion-Based Onbody Activity Recognition Systems. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 20–29).

Kunze, K., Lukowicz, P., Junker, H., & Tröster, G. (2005). Where am I: Recognizing On-body Positions of Wearable Sensors. *Location-and Context-Awareness*, *3479*, 264–275.

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity Recognition using Cell Phone Accelerometers. *ACM SigKDD Explorations Newsletter*, *12*(2), 74–82.

Lee, S. H., Park, H. D., Hong, S. Y., Lee, K. J., & Kim, Y. H. (2003). A Study on the Activity Classification Using a Triaxial Accelerometer. In *Proceedings of the 25th Annual International Conference of Engineering in Medicine and Biology Society* (pp. 2941–2943). IEEE.

Lester, J., Choudhury, T., Kern, N., Borriello, G., & Hannaford, B. (2005). A Hybrid Discriminative / Generative Approach for Modeling Human Activities. In *IJCAI, 2005* (pp. 766–772). Edinburgh.

Lockhart, J. W., & Weiss, G. M. (2014). Limitations with activity recognition methodology & data sets. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, 747–756.

Mathie, M. J., Celler, B. G., Lovell, N. H., & Coster, A. C. F. (2004). Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, *42*(5), 679–687.

Mathie, M. J., Coster, A. C. F., Lovell, N. H., & Celler, B. G. (2004, apr). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement*, *25*(2), R1–R20.

Maurer, U., Smailagic, A., Siewiorek, D. P., & Deisher, M. (2006). Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)* (pp. 113–116). IEEE.

Michahelles, F., & Schiele, B. (2005, jul). Sensing and Monitoring Professional Skiers. *IEEE Pervasive Computing*, *4*(3), 40–46.

Minnen, D., Starner, T., Essa, I., & Isbell, C. (2006). Discovering Characteristic Actions from On-Body Sensor Data. In *10th IEEE international symposium on Wearable computers* (pp. 11–18). IEEE.

Mizell, D. (2003). Using Gravity to Estimate Accelerometer Orientation. In *2012 16th International Symposium on Wearable Computers* (pp. 252–252). IEEE Computer Society.

Mueen, A., Keogh, E., Cash, S., & Westover, B. (2002). Exact Discovery of Time Series Motifs. In *SDM* (pp. 473–484).

Mueen, A., Keogh, E., & Young, N. (2011). Logical-Shapelets : An Expressive Primitive for Time Series Classification. In *17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1154–1162). ACM.

Oberg, T., Karsznia, a., & Oberg, K. (1993, jan). Basic gait parameters: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development*, *30*(2), 210–23.

Pascoe, J. (1998). Adding Generic Contextual Capabilities to Wearable Computers. In *2nd International Symposium on Wearable Computers* (Vol. 44, pp. 92–99). Pittsburgh, PA, USA: IEEE.

Patel, S., Mancinelli, C., Healey, J., Moy, M., & Bonato, P. (2009, jun). Using Wearable Sensors to Monitor Physical Activities of Patients with COPD: A Comparison of Classifier Performance. *2009 6th International Workshop on*

Wearable and Implantable Body Sensor Networks, 234–239.

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009, mar). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE transactions on bio-medical engineering*, *56*(3), 871–9.

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009, apr). Activity identification using body-mounted sensors–a review of classification techniques. *Physiological measurement*, *30*(4), R1–33.

Quinlan, R. J. (1993). *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers Inc.

Quwaider, M., & Biswas, S. (2008). Body posture identification using hidden Markov model with a wearable sensor network. *Proceedings of the ICST 3rd international conference on Body area network*(Article No. 19).

Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005). Activity Recognition from Accelerometer Data. In *AAAI-05 conference* (pp. 1541–1546). Pittsburgh, Pennsylvania.

Reiss, A. (2014). *Personalized Mobile Physical Activity Monitoring for Everyday Life* (Unpublished doctoral dissertation).

Reviews, B. F. T. (2014). *Activity tracker comparison chart.* Retrieved 29/5/2014, from http://www.bestfitnesstrackerreviews.com/comparison-chart.html

Saponas, T. S., Lester, J., Froehlich, J., Fogarty, J., & Landay, J. (2008). *iLearn on the iPhone : Real-Time Human Activity Classification on Commodity Mobile Phones* (Tech. Rep.). University of Washington, Dept. of Computer Science & Engineering.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. M. (2014). *Fusion of smartphone motion sensors for physical activity recognition* (Vol. 14) (No. 6).

Simoes, M. A. (2011). *Feasibility of Wearable Sensors to Determine Gait Parameters* (Unpublished doctoral dissertation). University of South Florida.

Song, K.-t., & Wang, Y.-q. (2005). Remote Activity Monitoring of the Elderly Using a Two-Axis Accelerometer. In *Proceedings of the CACS Automatic Control Conference* (pp. 18–19).

Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., & Tr, G. (2008, apr). Wearable Activity Tracking in Car Manufacturing. *Pervasive Computing, IEEE*, *7*(2), 42–50.

Sun, L., Zhang, D., Li, B., Guo, B., & Li, S. (2010). Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations. *Ubiquitous Intelligence and Computing*, 548–562.

Taylor, K., Abdulla, U. A., Helmer, R. J., Lee, J., & Blanchonette, I. (2011). Classification With Smart Phones For Sports Activities. In *the 5th Asia-Pacific Congress on Sports Technology (APCST)* (Vol. 00, pp. 428–433).

Tentori, M., & Favela, J. (2008, apr). Activity-Aware Computing for Healthcare. *Pervasive Computing, IEEE*, *7*(2), 51–57.

Wang, Y., Lin, J., Annavaram, M., Jacobson, Q. a., Hong, J., Krishnamachari, B., & Sadeh, N. (2009). A framework of energy efficient mobile sensing for automatic user state recognition. In *the 7th international conference on Mobile systems, applications, and services - MOBISYS '09* (pp. 179–192). New York, New York, USA: ACM Press.

Yang, J. (2009). Toward Physical Activity Diary : Motion Recognition Using Simple Acceleration Features with Mobile Phones. *Data Processing*, 1–9.

Ye, L., & Keogh, E. J. (2009). Time Series Shapelets : A New Primitive for Data Mining. In *the 15th ACM SIGKDD international conference.* ACM.

Yuwono, M., Su, S. W., Moulton, B. D., & Nguyen, H. T. (2012). Gait Cycle Spectrogram Analysis using a TorsoÂŋattached Inertial Sensor. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology*

Society (EMBC) (pp. 6539–6542). IEEE.

Zappi, P., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., & Tr, G. (2007). Activity recognition from on-body sensors by classifier fusion : sensor scalability and robustness. In *3rd International Conference on Intelligent Sensors, Sensor Networks and Information* (pp. 281–286). IEEE.