Dr. Droid: Assisting Stroke Rehabilitation Using Mobile Phones

Andrew Goodney[1], Jinho Jung[2], Scott Needham[1], and Sameera Poduri[1]

[1] University of Southern California, Los Angeles CA 90089, USA
{goodney,sneedham,sameera}@usc.edu,
http://robotics.usc.edu
[2] KJITC, Gyunggi-Do, Korea
visusee@mnd.go.kr

Abstract. In this paper we present our initial work on a mobile phone application for assisting stroke rehabilitation. We believe that using a mobile phone to administer and track stroke rehabilitation is novel. We call our system Dr. Droid and focus on the automated scoring of motions performed by patients being administered the Wolf Motor Function Test (WMFT) by placing a smart phone in a holster at the patients wrist. We have developed a complete software application that administers the test by giving audio and visual instructions. We collect a motion trace by sampling the 3-axis accelerometer available on the phone. We double-integrate the acceleration data and apply a novel reorientation algorithm to correct for mis-alignment of the accelerometer. Using dynamic time warping and hidden Markov models we assign an objective, quantitative score to the patient's exercises. We validate our method by performing experiments designed to simulate the motions of a stroke patient.

Keywords: mobile sensing, gesture recognition, personal health monitoring, telemedicine, stroke rehabilitation.

1 Introduction

Each year in the United States approximately 795,000 people suffer from a stroke. Of those strokes, about 29% are fatal [14]. Many of the remaining 71% of stroke patients will require some form of rehabilitation. Stoke rehabilitation is performed in the hope that the patients will maintain or regain range and detail of motion lost during the stroke.

Damage to the brain caused by a stroke often manifests itself as paralysis or loss of motor control on one side of the body, particularly with one limb. Several tests designed to quantify and monitor the rehabilitation progress of a stroke patient exist. In this work we focus on the Wolf Motor Function Test [11] (WMFT). The WMFT test focuses on range of motion in the upper extremity with a set of 17 exercises such as lifting a soda can to the mouth or turning a key. Table 1 shows a list of these exercises.

Scoring for the WMFT is based on two scores. The first score is the time required to perform the exercise. This time is measured by a test administrator

M. Griss and G. Yang (Eds.): MOBICASE 2010, LNICST 76, pp. 231–242, 2012.

Table 1. Wolf Motor Function Test task descriptions

1. Forearm to table (side)$^{\dagger(1)}$
2. Forearm to box (side)
3. Extend elbow (side)$^{\dagger(2)}$
4. Extend elbow (weight)
5. Hand to table (front)
6. Hand to box (front)
7. Weight to box*
8. Reach and retrieve
9. Lift can$^{\dagger(3)}$
10. Lift pencil
11. Lift paper clip
12. Stack checkers
13. Flip cards
14. Grip strength*
15. Turn key in lock$^{\dagger(4)}$
16. Fold towel
17. Lift basket

* Indicates un-timed tasked
† Indicates activities evaluated in this paper

using a stop-watch. The maximum possible score is 120 seconds. The second score is the functional analysis (FA) score. This score ranges from zero (meaning the affected limb can not perform the activity at all), to five (meaning the motion appears normal). The FA score is determined 'off-line' by a specially trained WMFT evaluator using a video of the patient.

A test that requires trained administers and evaluators greatly increases the cost of administering the WMFT and similar tests. Additionally, these tests are usually performed in an office or clinical setting under video observation. This setting maybe uncomfortable for the patient (which may lead to reluctance to attend future testing sessions), while adding additional overhead to the cost.

Researchers have shown that the variations between trained WMFT administrators and evaluators is low and the test reliable [9,4], at the cost of several hours of time for one or two trained practitioners. Additionally, timing errors of up to one second are likely when using a stop-watch to time the exercises [6]. This same work showed that automated timing of the WMFT is possible using wrist mounted accelerometers.

In this work we focus on both automating and scoring the WMFT in such a way as to provide quantitatively meaningful scores, while reducing the costs associated with the test and improving the patients experience.

Dr. Droid is our platform to assist stroke rehabilitation. We have devoped an application that runs on a smart phone worn on the patients wrist during the WMFT. Accelerometers in the device will record the motion of the patient. This acceleration data is then processed to give a quantitive score judging the quality of motion. For the hardware platform we use mobile phones running the Android operating system. Mobile phones are a good choice because they

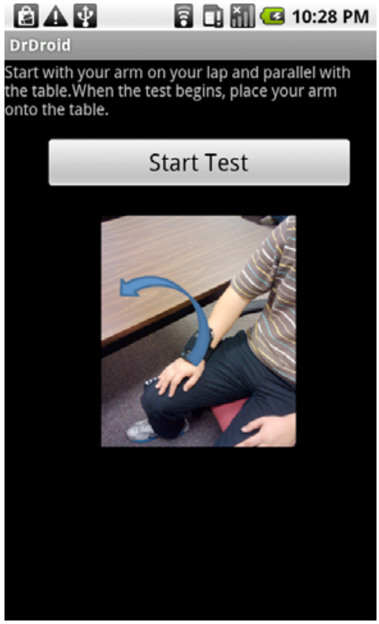


Fig. 1. Screen shot of the Dr. Droid application showing the wrist mounted phone and visual cues

are nearly ubiquitous and provide a rich array of sensor modalities at a cost approaching free. Additionally, the Android operating system provides an open, flexible platform that allows rapid application development. Combined together, these two aspects of the Android mobile phone allow us to provide an inexpensive way to administer the test at a time and location of the patients choosing, with at least as much quantitive information (time and FA score) as is currently available. Additionally, since the application is hosted on a smart phone, we enable a host of other telemedicine and digital health care applications.

There are four main contributions of this work: 1. We use a mobile phone in a novel way to make stroke rehabilitation inexpensive and comfortable for the patient, while also returning standardized, objective scores; 2. We develop two scoring algorithms based on a novel reorientation technique that corrects for errors caused by mis-alignments in the accelerometer sensor data. Thus we allow some flexibility in how the phone is mounted to the patients wrist; 3. We have implemented a complete system running on Android-based smart phones; 4. And finally we provide a quantitive comparison of three scoring algorithms under two testing modalities.

The rest of this paper is organized as follows: section 2 describes the Dr. Droid smart phone application and the scoring algorithms; section 3 describes the experiments conducted to verify our algorithms; section 4 describes related work in the literature; finally, we conclude and discuss the future work we aim to undertake in section 5.

2 Dr. Droid

This section describes the Dr. Droid application and scoring algorithms. First we provide a brief description of the system design and application operation. Second we describe in detail the algorithms we use to arrive at an objective score for a patients motion.

2.1 Assistive Application and Remote Control

The Android operating system provides a Java based programming framework for application development [1]. We used this framework to develop the Dr. Droid application. The Dr. Droid application provides several features. Users first may decide to perform an individual or a series of tests. Dr. Droid then prompts the user with visual, textual and spoken descriptions of the exercise under test. The user then performs the required action. Finally, our scoring algorithms (see section 2.2) compare the performed action to a template and return a set of scores. When performing a series of exercises the user is then presented and prompted with the next test. Finally, a database of past scores is retained, allowing the user to track the progress and report results to a practitioner. Given the average age of stroke patients and considering their ability to control a mobile device we have attempted to develop an as easy to use application as possible.

During development of Dr. Droid, we realized that it can be awkward to push buttons on the phones display while it is holstered at the wrist. As stroke patients in rehabilitation have limited mobility, we also developed a remote-control application using the XMPP protocol. This application also runs on Android phones and allows an assistant (medical practitioner or friend/family) to start and stop the timing of a test or to advance to the next test. XMPP is an open messaging protocol, thus our remote control application is not limited to running on a second phone. The remote control can be run on a desktop, laptop or tablet PC.

2.2 Data Processing and Scoring Algorithms

The core goal of Dr. Droid is to score how naturally a patient moves while performing an excercise. To do so, we perform an analysis that is very similar to gesture recognition. Gesture recognition is used in several related fields, from handwriting recognition on tablet-PCs or PDAs to gaming environments like the Nintendo Wii. However, our problem is slightly different. In gesture recognition the software system is trained with a ‘vocabulary’ of gestures and an unknown action is evaluated against the vocabulary and the most likely match is chosen as the detected gesture. In this work we know a-priori which gesture (exercise) is being performed, we only need to know how ‘well’ or ‘poorly’ it was performed.

Many machine learning algorithms have been used for gesture recognition, see the X-wand [19] project for an example where several were evaluated with the same hardware. In order to provide meaningful scores to the patient, we selected algorithms that provide a numeric metric of the similarity between two gestures.

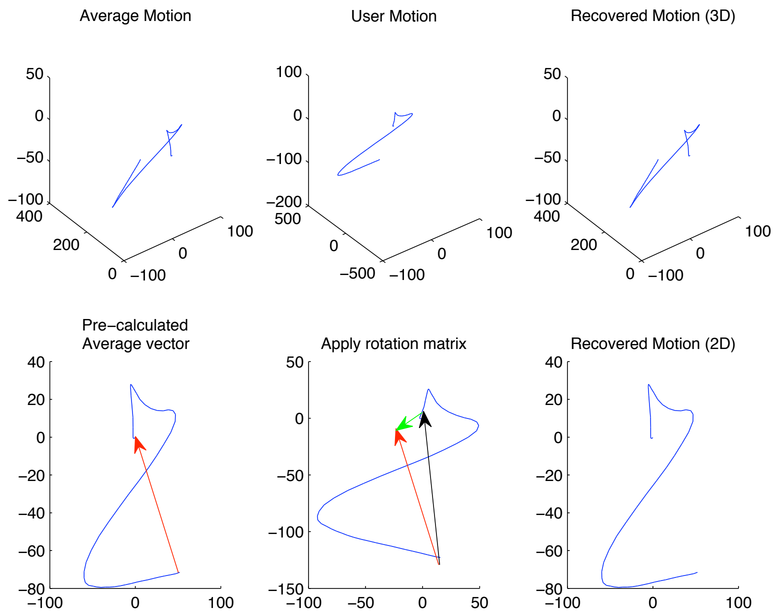


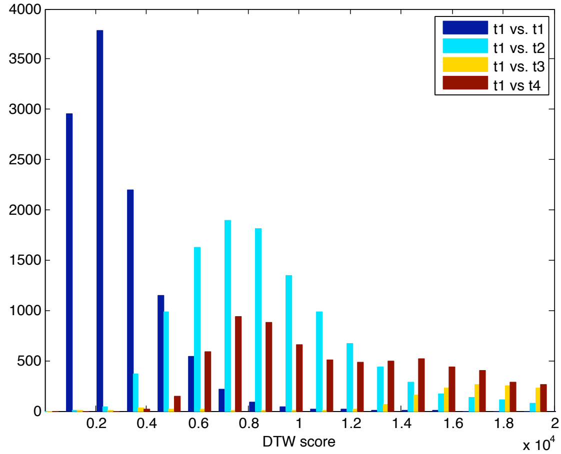
Fig. 2. Graphical illustration of our novel path-based reorientation algorithm

Two algorithms, dynamic time warping [7] and hidden Markov models [5] are a natural fit to this constraint. DTW is based on a dynamic programming technique and finds the minimal cost required to warp one signal onto the other, as long as the difference between two signals can be measured by some distance function. HMMs return the likelihood that a given signal belongs to a given (previously trained) class.

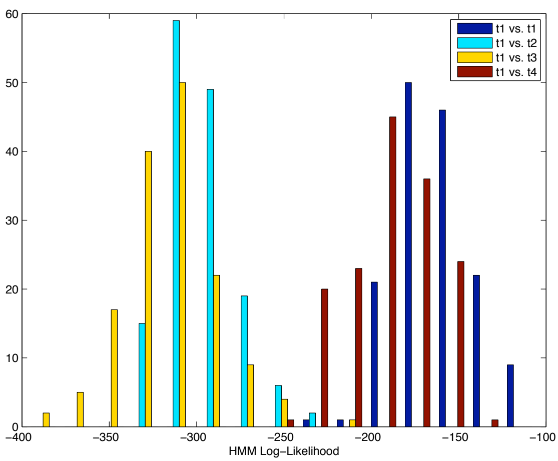
The Dr. Droid application collects acceleration data from the 3-axis accelerometer that is a standard feature of Android smart phones. The data is sampled at 50Hz. The WMFT allows a maximum test duration of 120 seconds, thus our acceleration traces contain no more than 18,000 samples. The phone used for testing, the Google Nexus One, has a 1GHz processor and is capable of scoring the motions using the above algorithms in a fraction of a second.

We include three scoring algorithms primarily because to date we have not had access to stroke patients to collect training data. Therefore, we cannot conclusively calculate which of our scoring methods correlates best with the FA scores given by trained observers. We also hypothesize that our algorithms may pick up on different aspects of a motions natural-ness and thus provide richer feedback about a patient's motion than a single FA score.

Path-Based Orientation Correction. Many difficulties with gesture recognition based on acceleration data are rooted in the orientation problem. If the orientation of the accelerometer is different between the training and test data,



(a) DTW discrimination



(b) HMM discrimination

Fig. 3. Histogram showing the ability of the DTW and HMM algorithm to score similar motions with similar scores. Intra-class scores (blue) cluster together, indicating similar scores are given to similar motions.

many gesture recognition algorithms preform poorly. Researchers have developed several techniques to reorient acceleration data [10], but this has proven to be difficult and many systems solve this problem by requiring that for optimal performance, the orientation of the device is constrained. Furthermore, variations in the amplitude of the action require some sort of normalization, which also reduces the accuracy of gesture recognition.

However, the prescribed nature of our testing environment leads to a novel solution for the above limitations. First, we observe that the amplitude of the test is fixed by the definition of the WMFT. Second, we observe that if one considers the starting point for a test as the origin of a 3-D reference frame, then the ending points for two motions performing the same exercise will have approximately the same coordinates. Acceleration data collected with a misoriented sensor will describe the path in a rotated reference frame.

Therefore, in order to reorient the motion traces to best align them with the template action, we need only to calculate the rotation necessary to align the end-point of a motion with the end point of the template. To do so we calculate the cumulative double integration of the acceleration data to obtain the 3-D path traced by the wrist mounted phone. Then we calculate the homogenous rotation matrix that will move the 3-D (x, y, z) coordinate at the end of the trace to the (x, y, z) coordinate of the template motion. The rotation matrix is then applied to every (x, y, z) coordinate in the path to reorient the data.

We found that using this reoriented path data provided the best results with the DTW and HMM algorithms. Please see figure 2 for a graphical illustration of the reorientation algorithm.

Dynamic Time Warping. Given two discrete signals $A[t]$ and $B[t]$ of length N and M respectively, DTW finds the best match between $A[t]$ and $B[t]$ by warping (compressing or stretching) $B[t]$ where necessary to minimize the distance between the signals. DTW is a flexible algorithm in that one may define the distance metric as necessary. We are working with the reoriented 3D path data, so signals $A[t]$ and $B[t]$ comprise a sequence of (x, y, z) points. We thus choose the distance metric to be the Euclidian distance between the points. Then $\forall i \,\epsilon\, 1..N$, $\forall j \,\epsilon\, 1..M$ we define the N by M *distance matrix*, D as follows:

$$D(i, j) =$$

$$\sqrt{(A[i]_x - B[j]_x)^2 + (A[i]_y - B[j]_y)^2 + (A[i]_z - B[j]_z)^2}$$

The DTW algorithm then finds the least-cost path through the distance matrix from $D(1, 1)$ to $D(N, M)$. Please see [7] for details of the algorithm.

An additional advantage of this algorithm is that it naturally handles sequences of different lengths. We expect the completion time to vary between tests, but do not need to normalize or resample the time series when using DTW. If $A[t]$ and $B[t]$ are the same signal, then the cost is zero. As the signals vary in path and time, the cost increases.

Thus, motions of similar path and time will receive lower scores. Less natural or longer motions will receive larger scores. Therefore, we conclude that the path based DTW score is a similarity metric and we use the cost calculated to traverse the matrix as the DTW component of the Dr. Droid score.

Hidden Markov Models. Hidden Markov models are a well known method for modeling noisy, time-varying processes. They are used in many areas such as speech, gesture and handwriting recognition; natural language processing, and

Table 2. Comparing the average results of our scoring algorithms across two emulation methods

(a) DTW Scoring

Exercise	Avg. Score - Normal	Avg. Score w/ weights	Avg. Score 'poor motion'
1	2079.35	4877.96	19041.85
2	5925.08	3834.49	10499.91
3	20313.27	17464.42	33212.71
4	15911.03	18382.09	160464.70

(b) HMM Scoring

Exercise	Avg. Score - Normal	Avg. Score w/ weights	Avg. Score 'poor motion'
1	-151.14	-184.08	-210.55
2	-300.47	-286.03	-376.95
3	-365.47	-378.73	-498.45
4	-225.32	-304.72	-377.48

(c) Smoothness Scoring

Exercise	Avg. Score - Normal	Avg. Score w/ weights	Avg. Score 'poor motion'
1	411.97	725.14	523.39
2	274.17	327.78	487.05
3	284.16	521.34	673.18
4	2456.56	5393.08	4590.48

bioinformatic applications such as DNA analysis. The process being modeled is assumed to be a Markov process where the state-sequence is unknown, but the emissions from the states are observed. Once trained, a HMM will return the probability that a given sequence of observed emissions was generated by the modeled process. We assume a linear topology for our Markov process.

In order to train the HMM, we need to express the 3D path as a sequence of discreet states. Similar to work done on 2D handwriting recognition [8], we propose the following approach that is novel in the 3D gesture processing literature. From one (x, y, z) point in the path data we discretize the direction of travel necessary to reach the next point. With three axes (x, y, z) and three possible directions {-,0,+} there are 27 possible movements. The series of 3D points is converted into sequence of states labeled with the corresponding state number. This state-sequence representation of the path data is then used in the training and querying of a HMM for each exercise in the WMFT. In order to determine the HMM component of the Dr. Droid score we calculate the log-likelyhood that a given path describes the given exercise.

Smoothness. If one observes a stroke patient performing the WMFT (see [3] for a video), it will be observed that the motion is often jerky or halting. In order to quantify the smoothness of a motion we have developed a third metric based on the average energy of the acceleration signal over time. For this metric we operate directly on the 3-axis accelerometer data. We segment the acceleration signal into frames of 12 samples with an overlap of 6 samples. The energy of the acceleration signal in each frame is calculated and the average of these energy values is returned as the Smoothness component of the Dr. Droid score.

3 Experimental Evaluation

In this section we describe two experiments undertaken to verify the correctness of our scoring algorithms. We first describe the experiments and then provide a quantitative comparison of the results.

The goal of the Dr. Droid scoring metrics can be summarized as follows: *similar motions should receive similar scores*. We first collected training data of natural motion. We chose four exercises from the WMFT (marked with a † in table 1) that are representative of the different types of motion in the WMFT. We called these four exercises 'task 1' through 'task 4.'

Three of the authors then performed the exercises 50 times with the phone holstered at the wrist (see figure 1). These training examples were used to develop the path based DTW algorithm and to train the HMM.

First, we needed to evaluate if the DTW and HMM algorithm would give similar (low) scores to similar motions. To do so we scored all of the training examples against each other, both intra-class and inter-class. See figure 2.2 for a histogram of the resulting scores for the DTW and HMM algorithm with task 1 used intra-class. In these figures we see that indeed the intra-class scores cluster at the lower magnitude end of each graph (log likelihood scores are negative). This implies that our algorithms meet our goal of scoring similar motions with low scores. The results for the other tasks were similar.

Next we performed two experiments to emulate the motion of a stroke patient. In the first experiment we attached 10 pounds of weight to the forearm under test. In the second experiment we executed motions that emulated the difficulty and limited range of motion exhibited by patients as seen in videos of the WMFT. For the DTW algorithm we chose a 'golden example' from among our 150 training examples for each exercise. The golden example was defined as the training example with the minimum average score when compared to all other training examples of a particular task. The golden example is then used as the template signal for our DTW algorithm. For the HMM and Smoothness algorithms we calculated the score as described above.

Table 2 contains a summary of the experiment results. These results overall show that we have met our goal of giving larger scores to unnatural motion. Average scores highlighted in blue indicate scores where the scoring algorithm correctly scored the experimental motions as 'worse' than the training examples. Scores highlighted in red indicate where the scoring algorithm gave a 'better' score to the unnatural motion. These results, while not ideal, are worth discussing.

First we note that all of the red results occur in the weighted experiment. While performing the weighted experiment we noted that the weights made the exercises more difficult, but they did not necessarily make our movement unnatural. Moving the arm when weighted required very deliberate and thoughtful actions, which may have yielded efficient, correct motions.

However, when we look at the results for the ‘poor motion’ experiment, we see very good results. All of the algorithms score these purposely unnatural motions much higher than the training data. Therefore we can conclude that our scoring algorithms, when used on data featuring a range of motions from very unnatural to natural will give meaningful scores.

4 Related Work

As stroke rehabilitation is an important medical necessity, there is a large body of work in this area. Similar to our research is robotic assisted research. Fasoli et. al and Lo et. al [12],[15] describe robotic therapies that allow the scoring and tracking of patient progress. Closer still to our work is the AutoCITE [16] project. AutoCITE is a computer and specialized workstation deployed at the patients home. AutoCITE allows computerized testing, however it is expensive and once installed, fixed in location.

Specific to the WMFT test, Wade et. al [6] explore using a wrist mounted accelerometer to automate the timing of the WMFT. A recent masters thesis by Avinash Parnandi [2] builds on this work to attempt a similar goal of automating the FA scoring aspect of the WMFT using machine learning techniques, however the system requires a custom sensor board and wearable computer.

Accelerometer based gesture recognition is a well developed field, with Pylvanainen contributing a book chapter [18] to Springer’s LNCS. More recently, Liu et al. [13] developed a DTW based gesture recognition for mobile phones, while Prekopcsak [17] used HMM and support vector machines to do gesture recognition on Nokia smart phones. All of these methods operate directly on the acceleration data and thus require orientation constraints.

5 Conclusion and Future Work

In this paper we have presented a application that uses a mobile phone to improve stroke rehabilitation. Using a mobile phone in this manner is novel and provides a less expensive and more flexible rehabilitation process to patients and medical practitioners. We describe Dr. Droid, a complete system that administers the stroke rehabilitation protocol and provides quantitive scores of a patients movement. To generate these scores we present a novel path based reorientation algorithm that improves the performance of our DTW and HMM algorithm. We tested Dr. Droid experimentally to verify the system design and scoring algorithms provide a working user experience and meaningful scores. Finally, we discuss several opportunities for future research.

Dr. Droid to date has served as a proof-of-concept application to test the suitability of the Android mobile phone to this applicaiton.

To prove our scoring algorithms at a minimum provide the same information about a patient as the functional ability score, we need to test our algorithms against data collected from actual stroke patients over the course of their recovery. In order to obtain this data we first hope to expand on the work of a colleague at USC who has kinematic models based on data captured from stroke patients. These models are in the form of joint angle vs. time. The path taken by the wrist and the acceleration data can be recovered from these models. The timing and functional ability scores for these models are also available. If our algorithms prove successful scoring the motions, then we envision a study where practitioners at a stroke rehabilitation clinic solicit volunteers to wear a holstered phone during their WMFT exercises.

We also will work to improve the Dr. Droid application as well. We hope to collect feedback from practitioners and patients regarding the features and usability of the application. We also will improve the remote control application which we see as expanding into an interactive tool to help practitioners, patients and caregivers track the patient rehabilitation progress. We will include automatic timing algorithms to further improve the accuracy of the Dr. Droid scores.

Telemedicine and health monitoring are large and active fields of research. At this point in the development of Dr. Droid we have to date focused on the scoring algorithms. We recognize that health data is considered some of the most sensitive data one might collect. Any number of privacy schemes may be applied to our system and will be added at a later date. Additionally, Internet enabling the application to allow automatic uploading of scores to a patients doctor or therapist is also and obvious extension of our work. We would like to recognize that we have not ignored these aspects in designing Dr. Droid, but thus far they have been beyond the scope of our work.

Acknowledgments. The authors would also like to thank Dr. Eric Wade for his thoughts and guidance.

References

1. `http://www.android.com/`
2. `http://cres.usc.edu/pubdb_html/files_upload/670.pdf`
3. Wolf motor function test, youtube video,
   `http://www.youtube.com/watch?v=SlJk88Nd-ZM`
4. Assessing wolf motor function test as outcome measure for research in patients after stroke. Stroke 32(7), 1635–1639 (2001)
5. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus (2006)
6. Wade, E., Parnandi, A., Mataric, M.: Automated administration of the automated administration of the wolf motor function test for post-stroke assessment. In: ICST 4th International ICST Conference on Pervasive Computing Technologies for Healthcare 2010 (2010)

7. Ellis, D.: Dynamic time warp in matlab, http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/
8. Elmezain, M., et al.: A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In: ICPR, pp. 1–4 (2008)
9. Morris, D.M., et al.: The reliability of the wolf motor function test for assessing upper extremity function after stroke. Archives of Physical Medicine and Rehabilitation 82(6), 750–755 (2001)
10. Prashanth, M., et al.: Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: SenSys 2008: Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, pp. 323–336. ACM, New York (2008)
11. Wolf, S.L., et al.: Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. Experimental Neurology 104(2), 125–132 (1989)
12. Fasoli, S.E., Krebs, H.I., Stein, J., Frontera, W.R., Hogan, N.: Effects of robotic therapy on motor impairment and recovery in chronic stroke. Archives of Physical Medicine and Rehabilitation 84(4), 477–482 (2003)
13. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uwave: Accelerometer-based personalized gesture recognition and its applications. In: IEEE International Conference on Pervasive Computing and Communications, pp. 1–9 (2009)
14. Lloyd-Jones, D., Adams, R.J., Brown, T.M., et al.: Heart Disease and Stroke Statistics–2010 Update: A Report From the American Heart Association. Circulation 121(7), 46–215 (2010)
15. Lo, A.C., Guarino, P.D., Richards, L.G., et al.: Robot-Assisted Therapy for Long-Term Upper-Limb Impairment after Stroke. N. Engl. J. Med. 362(19), 1772–1783 (2010)
16. Lum, P.S., Taub, E., Schwandt, D., Postman, M., Hardin, P., Uswatte, G.: Automated constraint-induced therapy extension (autocite) for movement deficits after stroke. J. Rehabil. Res. Dev. 41(3A), 249–258 (2004)
17. Prekopcsak, Z.: Accelerometer based real-time gesture recognition. Poster Preview (2008)
18. Pylvänäinen, T.: Accelerometer Based Gesture Recognition Using Continuous HMMs. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3522, pp. 639–646. Springer, Heidelberg (2005)
19. Wilson, A., Shafer, S.: XWand: UI for intelligent spaces. In: CHI 2003: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 545–552. ACM, New York (2003)