Automatic detection of stereotyped hand flapping movements: two different approaches

Nuno Gonçalves, José L. Rodrigues, Sandra Costa, Filomena Soares, *Members, IEEE*

Abstract— Stereotypical motor movements are one of the most common and least understood behaviors occurring in individuals with Autism Spectrum Disorder (ASD). The traditional methods for recording the number of occurrences and duration of stereotypies are insufficient and time consuming. Thus the objective of this study is to automatically detect stereotypical motor movements in real time considering two different approaches. The first approach uses the Microsoft sensor Kinect and gesture recognition algorithms. The second approach uses a trademark device of Texas Instruments with built-in accelerometers and statistical methods to recognize stereotyped movements. The two proposed systems were tested in children with Autism Spectrum Disorders (ASD) and the results are compared. This study provides a valuable tool to monitor stereotypes in order to understand and to cope with this problematic. In the end, it facilitates the identification of relevant behavioral patterns when studying interaction skills in children with ASD.

Keywords - Stereotypical motor movements, Kinect Sensor, accelerometer, gesture recognition, ASD

I. INTRODUCTION

Autism Spectrum Disorders (ASD) affects 1 in 150 children [1] and typically manifests itself during the first three years of life, being defined as a global development disorder [2-5]. Stereotyped behaviors, typically performed by people with ASD, are defined as repetitive interests and/or motor or vocal sequences that appear to the observer to be invariant in form and without any obvious eliciting stimulus or adaptive function [6].

Typically, autistic children have stereotypical behavior in particular when there is a change in their daily routine or when the environment is full of new perceptions; as a consequence, they tend to block the excess of sensations, manifesting repetitive movements. Therapists measure this behavior using traditional methods such as paper-pencil rating scales, direct observation or video-based methods that normally are inaccurate and time consuming, depending on the observer.

The work presented in this paper is part of a collaborative project [7-9] between the University of Minho, APPACDM (an association for mentally disabled people) of Braga and a group of schools with Special Education units, where the main goal is to improve social life of children with ASD, in particular to promote their social interaction and communication, by using a robotic platform as mediator/reward.

Thus, the main objective of this work is to understand if the Kinect sensor and the eZ430-Chronos watch with the accelerometers can be used as a time-efficient tool to automatic detect in real-time stereotypical behaviors of children with ASD. This will allow characterizing children's behavior during interventions sessions, when they are exposed to new situations.

The Kinect sensor is based on the concept of Human-Device interaction through the user's senses, especially hearing and vision. The sensor shown in Figure 1 contains a RGB color camera, a depth sensor and multi microphone array which allows capturing body motion in 3D, gesture and face recognition, obtaining the users joints coordinates. The depth sensor is an infrared laser in combination with a CMOS monochrome sensor which enables the Kinect device processing 3D scenarios in any ambient lighting.

Figure 1 – Kinect Sensor

The eZ430-Chronos from Texas Instruments, Figure 2, is a highly integrated, wireless development system inside a smart watch. Chronos is a reference platform for many applications, such as wireless sensor nodes for remote data collection and other applications. The integrated wireless feature allows the Chronos to act as a central hub for nearby wireless sensors such as pedometers and heart rate monitors [11].

This paper is structured as follows: Section I introduces the research focus of the paper and Section II describes the

Nuno Gonçalves is a Researcher in University of Minho. He is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (email: nuno.goncalves@algoritmi.uminho.pt)

José L. Rodrigues is a Researcher in University of Minho. He is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (email: jose.rodrigues@algoritmi.uminho.pt)

Sandra Costa is a PhD Student in University of Minho (SFRH/BD/71600/2010). She is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (e-mail: sandra.costa@algoritmi.uminho.pt).

Filomena Soares is with University of Minho. She is with the R&D Centre Algoritmi, Campus de Azurém, 4800-058 Guimarães, Portugal (e-mail: filomena.soares@algoritmi.uminho.pt)

related work. Section III shows the applied methodology. Section IV refers to the developed applications developed. Results are presented and discussed in Section V. The paper ends with conclusions and the future steps to improve the applications.

Figure 2 - eZ430-Chronos watch [11].

II. RELATED WORK

Recently, systems have been developed to automatically detect stereotypical behaviors using accelerometers and pattern recognition algorithms with promising results.

The work presented by [12, 13] uses a three-axis accelerometer to acquired data. The accelerometer is placed on each wrist and chest of the child to detect hand flapping and body rocking movement associated to these problematic, Figure 3.

Figure 3 – (a) A wireless accelerometer placed on each wrist. (b) A wireless accelerometer placed on the chest. (c) Receiver for sensor data. (d) An image of a child in the laboratory setting. (e) The video coding software that allows frame-accurate annotation [12].

The pattern recognition algorithm used is based on the attainment of the characteristics at each component of the acceleration (X, Y, Z axis), where the distance between the average for each component of acceleration is analyzed to know the orientation of the sensor. The variance of the movements of different directions, the correlation coefficient to capture simultaneous movements and the entropy to identify the types of stereotyped movements are also considered. These characteristics were analyzed using the Weka software [14], which carries out the cross-validation with the input signals with the analyzed signals.

Min Cheol-hong *et al.* [15] also used three-axis accelerometer to acquire data. The data was analyzed using time-frequency and observing frequency band powers. Also Linear Predictive Coding (LPC) method was used for classification of the stereotypic and self-injurious behaviors.

The Kinect sensor allows capturing movement in 3-D, gesture recognition, voice command recognition, face detection and robotics [16-19, 22], but there isn't no related work to automatically detect stereotyped movement.

III. METHODOLOGY

In this section the methodology to recognize stereotypical movements with the Kinect sensor and the eZ430-Chronos watch is described.

A. Dynamic time warping

When using the Kinect sensor the first step is to extract the coordinates of the selected joints: the joints of the hands, wrists and head. Since the coordinates of the selected joints are referenced to the sensor, ie the Kinect sensor is O(0,0,0) and the gesture recognition algorithm does not have to depend on the user's height or where he/ she is sitting, it is necessary convert the user as a new reference. For convenience it was chosen the shoulders of the user, as the new reference coordinate O(0,0,0), Figure 4.
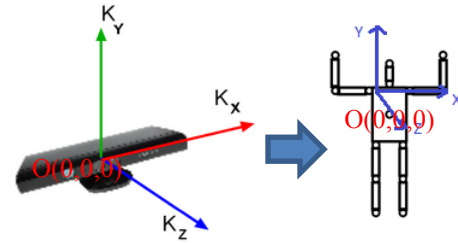

Figure 4 – Illustration of the converted reference

To recognize stereotypical motor movement with the Kinect sensor it was used the Dynamic Time Warping (DTW) algorithm [19-22]. This algorithm is based on a dynamic programming that measures the similarity between patterns which may vary with time of different durations. The main concept of the algorithm is to compare characteristics of a particular pattern to be checked against a template reference previously registered.

The algorithm has been used in voice recognition patterns, signature verification or gesture recognition [23, 25].

The final objective of the DTW algorithm is to compare a pattern standard with a sample input pattern. The algorithm has to yield a match result between the two patterns. In the end, a measure of overall distance/similarity between the two patterns is generated. The smaller the distance, more similar the patterns will be.

Assuming that we have two time sequences, X and Y, of length n and m respectively, where $X = \{x_1, ..., x_n\}$ and $Y = \{y_1, ..., y_m\}$ and the goal is to compute the matching cost matrix: *DTW(X,Y)*.

The two sequences can be arranged on the sides of a matrix, with one sequence placed on the top and the other is located up to the left hand side, Figure 5; both sequences start on the bottom left side of the matrix.

To find the best match or alignment in the matching cost matrix, it is necessary to find a path, also known as the warping path, through the grid which minimizes the total

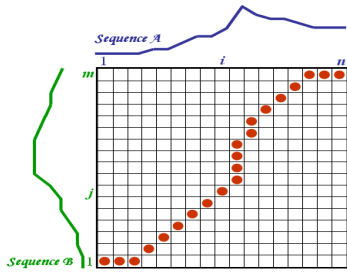distance. The procedure involves finding all possible routes through the matrix.

Figure 5 – Example of the matching cost matrix [20]

The major optimizations to the DTW algorithm arise from the necessity of a good path through the grid. These were outlined in Sakoe and Chiba [23] and can be summarized as follows:

1. Boundary condition: the path has to start at the bottom left and ends at the top right;

2. Monotonic condition: the path cannot turn back on itself; both i and j indexes either stay the same or increase, they never decrease;

3. Continuity condition: the path advances one step at time. Both i and j can only increase by at most 1 on each step along the path;

4. Slope constraint condition: the path should not be too steep or too shallow. This prevents very short sequences matching very long ones.

Figure 6, shows an example of an admissible warping path satisfying the above conditions.
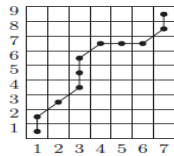

Figure 6 – Example of the warping path [23]

The matching cost is computed based on dynamic programming presented as:

$$D(i,j) = d(x_i,j_i) + min \begin{cases} D(i,j-i) \\ D(i-1,j) \\ D(i-1,j-1) \end{cases} \quad (1)$$

where the distance function $d(x_i, y_i)$, also known as the Euclidian Distance, is defined as

$$d(x_i,j_i) = \sqrt{(x_i - y_{i)}{}^2} \quad (2)$$

where x_i and y_i are the coordinates of point i

After computing the matching cost matrix the DTW distance is calculated by dividing the sum of the warping path by the length of the standard pattern.

B. Algorithm based on statistical methods

In order to recognize stereotypical motor movement with the eZ430-Chronos watch it is necessary to characterize the signal acquired from a previously expressed stereotyped movement. Several experimental tests were performed with children with ASD in order to characterize the signal in terms of the mean values of Standard Variation, Root Mean Square (RMS) and Peaks and Valleys. Dedicated software was developed using statistical methods based on the parameters previously referred.

The standard deviation is used to recognize the beginning of the movement; when it is near zero indicates no movement. Thus when the value of the standard deviation is equal or greater than the expected average is considered a movement of interest, Figure 7.
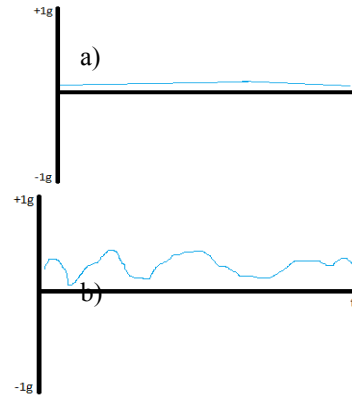

Figure 7 – a) no movement detected; b) movement detected

If the motion is considered of interest, it is necessary to check the signal energy of the movement, which can be characterized by the value of RMS. The movement abrupt acceleration corresponds to high energy signals matching to higher RMS values. If the RMS value is equal or greater than the expected average value (calculated through the stereotypes movements previously recorded), the gesture remains as a movement of interest.

Although the movement could be considered of interest, it may not be considered as a stereotypical gesture. It is necessary to get some more features such as the number of peaks (it indicate signal inflection when the acceleration is greater than zero) and valleys (indicates how many times the wave crossed zero). So when the number of peaks and valleys is equal or greater than the expected average, the gesture is considered a stereotypy.

IV. SYSTEMS DEVELOPED

In order to use the Kinect sensor and the eZ430-Chronos watch, it was necessary to develop interface applications. In this section the applications are described.

A. Application developed with the Kinect Sensor

An application based on WPF (Windows Presentation Form) and C# language was developed. After the application is initialized, it is necessary to wait until the Kinect detects a user (represented by the red color in the depth image, Figure 8). When the user is detected, it has to record the hand flapping movement to serve as a template sequence to the

DTW algorithm. By pressing the *Hand flapping* button, the gesture is recorded for twenty frames. The value of twenty frames was chosen, after doing some tests, being as a sufficient value for a good template sequence. When the recording ends the application goes to the reading mode, ready to identify the stereotypes as shown in Figure 8
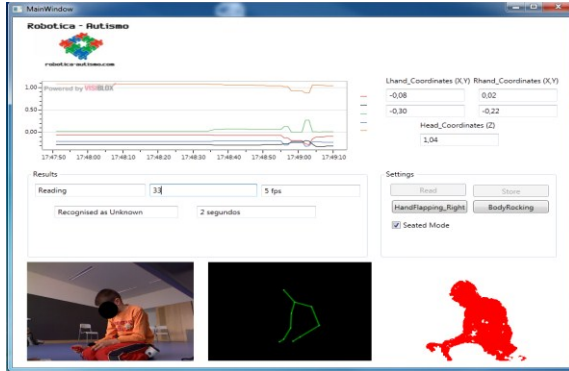

Figure 8 –Application developed for the Kinect sensor

The sampling period (sample sequence) chosen for the application in the reading mode was five frames. This value was chosen by trial and error. The application at every five frames stores the hand coordinates in an array. Since the hand flapping movement it is generally performed along the Oy axis, only the Oy coordinates were considered relevant. Through the DTW algorithm it is verified if there is a similarity between the template sequence and the array with the hand coordinates. If a match is found, it is considered a stereotypical movement and starts to count the time duration of the hand flapping.

B. Application developed with the eZ430-Chronos

For the use of the eZ430-Chronos watch, it was developed an interface application using the LabVIEW from National Instruments.

To receive data from the accelerometer for further analysis, it was necessary to establish the connection protocol between the module USB-RF Access Point and LabVIEW. The signal from the eZ430-Chronos watch is a digital signal which varies from 0 to 1024 bits. This signal has been subjected to analogue conversion in order to see on a wave Chart the characteristic signal of a stereotyped motion. So, digital values were converted to analogue values, which vary between -1 g to +1g (g represents the unit measure of gravity acceleration). Subsequently, the signal is decomposed into scalar components (acceleration axis Ox, Oy and Oz) allowing simple the characterization of the analyzed acceleration.

The developed program collects forty sample of the acceleration signal, in one second sampling [13]. The program automatically runs the statistical analysis module in order to identify the pattern behavior. If this sample is within the predefined threshold regarding each statistical parameter explained in Section III B), it is considered a stereotyped movement, Figure 9.
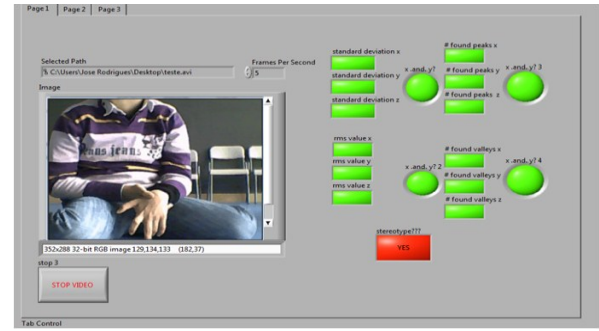

Figure 9 - Stereotype identification with the eZ430-Chronos and the LabVIEW application

V. RESULTS

The sessions were performed with children with autism and took place in two different schools of Leça de Palmeira (Portugal) in the Special Education Unit Department. The systems were tested on four ASD children with average age of M = 10.6 years and standard deviation SD = 1.95 in two intervention sessions of ten minutes. All sessions were recorded for further video analysis to compare with the obtained results.

Considering that the eZ430-Chronos watch is placed on the right wrist, for this study it was only considered the hand flapping of the right arm as a stereotyped behavior.

Table 1 shows the obtained results regarding the quantification of the stereotyped motor movements. The V_R represents the number of times that the child manifested the hand flapping movement. These values were obtained after analyzing the recorded videos of the sessions. The number of times that the Kinect sensor has detected a stereotyped movement is representing by K_R. The ACC_R shows the results obtained with the eZ430-Chronos watch.

Table 1 – Results obtained by the Kinect sensor and the eZ430-Chronos.

Child	V_R	K_R	ACC_R
1	16	38	11
2	17	6	20
3	0	0	0
4	1	1	7
Average	8,50	11,25	9,50
Standard Deviation	9,26	18,03	8,35

It can be observed in Table 1, the number of stereotypical movements detect by the Kinect sensor in the first child were higher than those performed by the child. This result demonstrates that the Kinect sensor obtained false-positive results. In the second child the Kinect sensor detected less stereotypical movements than the stereotypes held by the child while the numbers of stereotypies detected by the eZ430-Chronos watch were higher.

Tables 2 and 3 show the results obtained with the Kinect sensor and eZ430-Chronos watch, respectively, comparing

the total real movements performed by the children to the corresponding movements detected by the developed applications. The comparison is performed in terms of:

- Total number of times that the children performed stereotyped movement (**Results Video Analyzed**);

- Total number of times that both applications detected stereotyped behavior (**Total Kinect detected** and **Total watch detected**)

- Total number of times that both applications detected stereotyped behavior that corresponds to the real stereotyped behavior performed by the children (**Total Kinect real detected** and **Total watch real detected**)

- Total number of times that both applications do not detected stereotyped behavior comparing to the stereotyped movement observed(**Total Kinect not detected** and **Total watch not detected**)

- Total number of false-positives detected by the both systems (**Total false positives Kinect detected** and **Total false positives watch detected**)

Table 2 – Results obtained by the Kinect

	Results Video Analyzed	Total Kinect Detected	Total Kinect real detected	Total Kinect not detected	Total false positives Kinect detected
Child 1	16	38	16	0	22
Child 2	17	6	6	11	0
Child 3	0	0	0	0	0
Child 4	1	1	1	0	0
Average	8,5	11,25	5,75	2,75	5,5
SD	9,26	18,03	7,32	5,5	11
Percentage	--	--	51%	32%	49%

Table 3 – Results obtained the eZ430-Chronos watch

	Results Video Analyzed	Total watch detected	Total watch real detected	Total watch not detected	Total false positives watch detected
Child 1	16	11	11	5	0
Child 2	17	20	17	0	3
Child 3	0	0	0	0	0
Child 4	1	7	1	0	6
Average	8,5	9,5	7,25	1,25	2,25
SD	9,26	8,34	8,18	2,5	2,87
Percentage	--	--	76%	15%	24%

It can be observed in Table 2, that the Kinect sensor detected 51% number of times of the stereotypical movement. It also can be observed that 32% number of times the Kinect did not detect the stereotypical movement performed by the children. This could be explained by the fact that the children left the field of view of the Kinect sensor. The 49% number of time that the Kinect sensor detected false-positives was due to the intervention of the investigator by the children, because the Kinect sensor cannot distinguish between the child and the investigator.

In Table 3, it can be observed that the eZ430-Chronos watch detected 76% number of times of the stereotypical movement. 15% number of times the eZ430-Chronos watch did not detected the stereotypical movement performed by the children. This result could be explained by the fact that one of the children tried to take of the eZ430-Chronos bracelet. It also can be observed that 24% number of times the eZ430-Chronos watch detected false-positives.

The results obtained with eZ430-Chronos showed to be more accurate, but has the disadvantage to worn the sensors on the wrists, a situation that proven to be not well accepted by the some of the children.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a new system based on the Kinect sensor and the eZ430-Chronos watch to automatic detect stereotyped behaviors, in particular the hand flapping. We applied the DTW algorithm to recognize the hand flapping gesture with the Kinect sensor. An algorithm based on statistical methods was developed to characterize hand flapping movement using the eZ430-Chronos. The systems were tested in a classroom environment in children with ASD.

The Kinect sensor detected 51% number of times the stereotypical movements, while the eZ430-Chronos watch detected 76% number of times.

Although the results showed to be promising, some aspects need to be improved. The noise of the depth image obtained by the Kinect sensor that can lead to false-positives in the identification could be suppressed by applying Image processing techniques. It is necessary to improve the accuracy of the application when the user sits at extremes distance (more than 3.5 meters), or too close to the Kinect sensor.

It is necessary to improve the algorithm used to recognize the hand flapping movement with the eZ430-Chronos, due to the false-positives detected. Also, it will be necessary to integrate the watch on a friendly bracelet to the user, and also the video synchronization needs to be improved.

It is worth mention that although both systems could work independently, for better accuracy and reliability of the results both methods should be complementary.

As future work, it is intended to improve the gesture recognition algorithms, the use of the Hidden Markov Models (HMM) could be a possibility and add the body rocking gesture recognition in both systems.

REFERENCES

[1]   U.S. CDC, "Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities.

[2]   Schopler, E., G. B. Mesibow (Eds), (1987). Neurobiological issues in autism, en Current issues in autism (series). New York: Plenum Press H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[3]   Giellberg, C. & M. Coleman (Eds) (1992). The biology of the autistic syndromes, 2nd ed.,Londres: Mackeith.

[4]   Happé, F. (1994), Introducción al autismo. Trad. María Núñez Bernardos (1998). Madrid:Alianza Editorial, S.A.

[5]   A. P. A., Diagnostic and Statistical Manual on Mental Disorders, 4th Ed., vol. IV-TR. Washington, DC: Amer. Psychiatric Publishing, 2000.

[6]   A. Baumeister and R. Forehand, "Stereotyped acts," in Int'l. Rev. of Res. in Mental Retardation: VI., Ed. New York, 1973, pp. 55-96.

[7]   Sandra Costa, Jorge Resende, Filomena Oliveira Soares, Manuel João Ferreira, Cristina P. Santos, Fátima Moreira, Applications of simple robots to encourage social receptiveness of adolescents with autism, in EMBC 2009, Engineering in Medicine and Biology Conference, September 2009, Minneapolis, Minnesota, EUA.

[8]   Sandra Costa, Cristina Santos, Filomena Soares, Manuel João Ferreira e Fátima Moreira, Promoting the interaction amongst autistic adolescents using robots, in EMBC 2010, Engineering in Medicine and Biology Conference, September 2010, Buenos Aires, Argentina.

[9]   Sandra Costa, Filomena Soares, Cristina Santos, Manuel J. Ferreira, Fátima Moreira, Ana Paula Pereira, Fernanda Cunha, An approach to promote social and communication behaviors in children with Autism Spectrum Disorders: Robot based intervention, in Ro-Man 2011, 20th IEEE International Symposium on Robot and Human Interactive Communication, August 2011, Atlanta, Georgia, EUA.

[10] Valentino Frati, Domenico Prattichizzo – "Using Kinect for Hand Tracking and rendering in wearable haptics" – IEEE World Haptics Conference 2011, Istanbul, Turkey

[11] http://processors.wiki.ti.com/index.php/EZ430-Chronos (accessed July 2012)

[12] Fahd Albinali, Mathew S. Goodwin, Stephen S.Intille, "Recognizing Stereotypical Motor Movements in the Laboratory and Classroom: A case Study with Children on the Autism Spectrum" ,in Ubicomp 2009, September 2009, Orlando, Florida, USA.

[13] Fahd Albinali, Mathew S. Goodwin, Stephen S.Intille, Wayne F. Velicer, "Automated Detection of Stereotypical Motor Movements", Journal of Autism and Developmental Disorders, Springer Ed, Netherlands, Vol. 41, Issue 6, pp. 770-782.

[14] http://www.cs.waikato.ac.nz/ml/weka/ (accessed in May 2012)

[15] Cheol-Hong, Ahmed H. Tewfik, "Automatic Characterization and Detection of Behavioral Patterns using Linear Predictive Coding of Accelerometer Sensor Data", in IEEE EMBS 2010, Engenineering in Medicine and Biology Society, September 2010, Buenos Aires, Argentina.

[16] Panjawee Rakprayoon, Miti Ruchanurucks, Ada Coundoul, "Kinect-based Obstacle Detection for Manipulator", in System Integration (SII), 2011 IEEE/SICE International Symposium on 2011, 20-22 December 2011, Kioto, Japan.

[17] K. K. Biswas, Saurav Kumar Basu, "Gesture Recognition using Microsoft Kinect®", in Proceedings of the 5th International Conference on Automation, Robotics and Applications, , Dec 6-8 2011, , Wellington, New Zealand.

[18] Jing Tong, Jin Zhou, Ligang Liu, Zhigeng Pan, and Hao Yan, "Scanning 3D Full Human Bodies Using Kinects" in IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 4, APRIL 2012.

[19] Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, Samuel Sonning, Sabina Sonning, "3D Head Pose Estimation Using the Kinect", in International Conference on Wireless Communications and Signal Processing (WCSP), 2011, 9-11 November 2011, Nanjing, China.

[20] http://www.psb.ugent.be/cbd/papers/gentxwarper/DTW algorithm.htm (accessed in July 2012)

[21] Müller, Meinard "Information Retrieval for Music and Motion", New York: Springer-Verlag, 1985, chapter 4

[22] http://kinectdtw.codeplex.com/ (accessed July 2012)

[23] Sakoe and Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-26, NO. 1, February 1978.

[24] Cory Myers, Lawrence R. Rabiner and Aaron E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, Volume 28, Issue 6, 1980, pp: 623 – 635.