BioMedical Engineering OnLine

Open Access

CrossMark

Test–retest reliability of Kinect's measurements for the evaluation of upper body recovery of stroke patients

A Mobini, S Behzadipour* and M Saadat

*Correspondence:
behzadipour@sharif.edu
Mechanical Engineering
Department, Sharif University
of Technology, Tehran, Iran

Abstract

Background: Performance indices provide quantitative measures for the quality of motion, and therefore, assist in analyzing and monitoring patients' progress. Measurement of performance indices requires costly devices, such as motion capture systems. Recent developments of sensors for game controllers, such as Microsoft Kinect, have motivated many researchers to develop affordable systems for performance measurement applicable to home and clinical care. In this work, the capability of Kinect in finding motion performance indices was assessed by analyzing intra-session and inter-session test–retest reliability.

Method: Eighteen stroke patients and twelve healthy subjects participated in this investigation. The intra-session and inter-session reliability of eight performance indices, namely mean velocity (*MV*), normalized mean speed (*NMS*), normalized speed peaks (*NSP*), logarithm of dimensionless jerk (*LJ*), curvature (*C*), spectral arc length (*SAL*), shoulder angle (*SA*), and elbow angle (*EA*), were assessed using intra-class correlation coefficient (*ICC*), standard error of measurement (*SEM*) and coefficient of variation (*CV*).

Results: The results showed that, among the performance indices, *MV*, *LJ*, C, SA and *EA* have more than 0.9 *ICC* together with an acceptable *SEM* and *CV* in both stroke patients and healthy subjects. Comparing the results of different therapy sessions showed that *MV*, *LJ* and C are more sensitive than other indices, and hence, more capable of reflecting the progress of a patient during the rehabilitation process.

Conclusion: The results of this study shows acceptable reliability and sensitivity across the sessions for *MV*, *LJ* and C measured by Kinect for both healthy subjects and stroke patients. The results are promising for the development of home-based rehabilitation systems, which can analyze patient's movements using Kinect as an affordable motion capture sensor.

Keywords: Reliability, Kinect, Stroke, Motor assessment, Neurorehabilitation

Background

Using virtual reality systems for stroke rehabilitation is a flourishing field in physical and neurological rehabilitation. Such systems can help patients have a more intensive and entertaining training. They are commonly composed of a sensory device to capture the patient's movements, and a computer interface to communicate with the patient and

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 2 of 13

guide him through the intended tasks. Various sensors have been used in the development of such systems, including very simple ones, such as IMUs (inertial measurement units), and more sophisticated ones, such as sensorized robotic arms. The potentials of such systems in physical and neurological rehabilitation have been investigated by various researchers. For instance, Tsoupikova et al. [1] and Wade et al. [2] have proven the capability of virtual reality systems in the hand rehabilitation of post stroke patients. Deutsch et al. [3] have investigated the efficacy of Nintendo Wii in the rehabilitation of cerebral palsy patients. Moreover, Medrum et al. [4] have studied the applicability of Nintendo Wii in the balance rehabilitation of neurologic patients. Patel et al. [5] have reviewed the utilization of wearable sensors in rehabilitation. Also, Bao et al. [6] have discussed the role Kinect can play in the recovery of upper limbs of stroke patients.

High cost of such systems, which is mostly due to the sensor hardware, is one of the main challenges. As a result, using modern motion sensing game controllers such as Microsoft Kinect [7], Nintendo Wii remote, and Balance Board [8] has been widely investigated. These new sensors have much lower cost than the clinical ones, namely Vicon [9] and Optotrak [10], because they are simpler, less accurate, and are being mass produced in over 20 million numbers. Despite being recently introduced to the market, these new motion sensors have been used in various commercial rehabilitation applications such as Wiifit [11], SeeMe [12], Virtualrehab [13], and JINTRONIX [14].

In several previous works, it has been shown that stroke patients' current state and amount of recovery can be assessed based on their hand movement quality. Rohrer et al. [15] have assessed movement smoothness changes during the recovery of stroke patients. Van Dokkum et al. [16] have investigated the contributions of kinematics in the assessment of upper limb motor recovery of stroke patients. Osu et al. [17] have studied the quality of hand movements using three dimensional curvature. Moreover, Balasubramanian et al. [18] have defined a new metric for quantifying the smoothness of hand movements of stroke patients. Hogan et al. [19] have discussed sensitivity of jerk smoothness measure in determining hand movement performance. These quality measures, which mostly include kinematic indices of motion such as mean velocity, dimensionless jerk, motion curvature, etc., can be computed automatically during therapy sessions in order to provide in-depth information about the patient's impairments and recovery process. Determining these features by using clinical measurement systems such as Xsens [20], Optotrak [10] and Vicon [9] is very expensive and can be applied only in labs equipped for this purpose. On the other hand, using game controllers, despite being more affordable and available in almost any environment, has drawbacks such as lower accuracy, precision, and, more importantly, the reliability of outputs.

In the related researches, validity and reliability of Wii balance board using Nintendo Wii Fit balance scores [21], and balance scores measured by a balance assessment software have been discussed [22]. Wii remote controller's accuracy, validity and reliability for measuring head posture [23] has also been determined. Kinect's ability in pose estimation [24] and measuring joints positions [25] has been investigated in the literature as well. Also, the reliability and accuracy of measuring the range of motion [26], joints positions [27], and joints angles [28] using Microsoft Kinect has been studied by various researchers.

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 3 of 13

Although many studies have been carried out on the accuracy and reliability of these sensors, the measurement targets have been the raw data which is directly measured by the sensors. Measuring performance indices and the corresponding measurement reliability, which is essential for application of such sensors in rehabilitation, have been mostly overlooked. The study of Elgendi et al. [29], in which they calculated speed and classified subjects using Kinect, is the closest one to using Kinect for measuring movement performance indices. However, even in this work the reliability of such measurements is not determined.

In order to fill this gap, this study attempts to assess the test–retest reliability of Kinect's measurements of performance induces for the evaluation of upper body recovery in stroke patients. Among various methods available for determining test–retest reliability, this work uses the intra-session and inter-session approach. Movement performance indices were measured on 30 subjects using Kinect and the measurement reliability was analyzed. Furthermore, the sensitivity of the performance indices measured by Kinect was investigated.
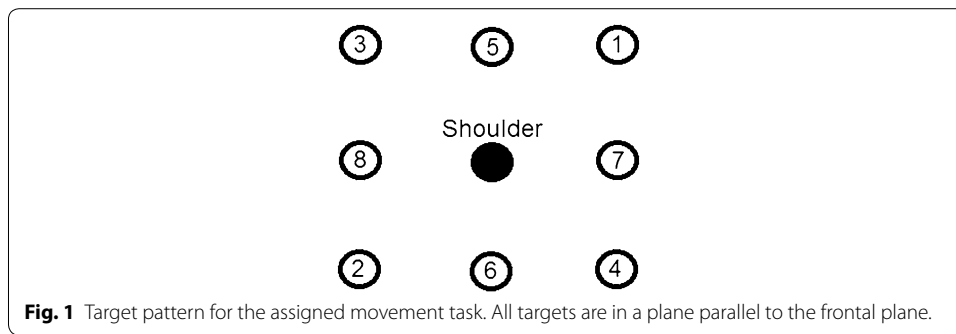
Methods

Subjects

A group of 18 stroke patients (8 females and 10 males; aged 50 ± 16) and 12 healthy subjects (4 females and 8 males; aged 46 ± 15) participated in this research. The tests were taken at the Red Crescent Society Rehabilitation Center (Tehran, Iran). Eleven of the patients were in the sub-acute phase (less than 6 months had passed from their unilateral cerebrovascular accident or CVA), and seven of them were at the chronic phase (more than 6 months had passed from their unilateral CVA). The inclusion criteria consisted of the existence of a single unilateral CVA and the occurrence of movements with more than 15 degrees in the impaired shoulder and elbow. However, any sever visual impairments, apraxia, or neglect syndromes lead to the exclusion of subjects from the tests. All subjects were informed about the study and willingly participated in the tests. Their consents were approved by the local scientific and ethics committees.

Data capture program

The subjects' hand movements were measured by the Microsoft Kinect for Xbox 360 using Microsoft Kinect's skeleton tracking driver version 1.7 [7]. A program was designed for this purpose using C# and Microsoft XNA game studio [30]. During the test, subjects sat or stood in front of a video screen running a graphical interface program. Kinect's distance from each subject was about 2.7 m. This was set so that Kinect could see the full body of the subject. As the sitting option was used in the skeleton tracking program and only upper body joints were tracked, there was no difference between tracking subjects in standing or sitting positions.

In the program, they were instructed to move their hands in order to intercept and catch several approaching balls. Balls were sent toward the subject using a predefined pattern shown in Fig. 1. All of the targets are on a plane parallel to the frontal plane but reaching them requires movements in the three dimensional space. The program also provided audio feedback to the patients based on their performance. During the tests, all upper body joints' positions (Hand, Wrist, Elbow, Shoulder, Shoulder Center, Head

Mobini *et al. BioMed Eng OnLine (2015) 14:75*

Page 4 of 13

Fig. 1 Target pattern for the assigned movement task. All targets are in a plane parallel to the frontal plane.
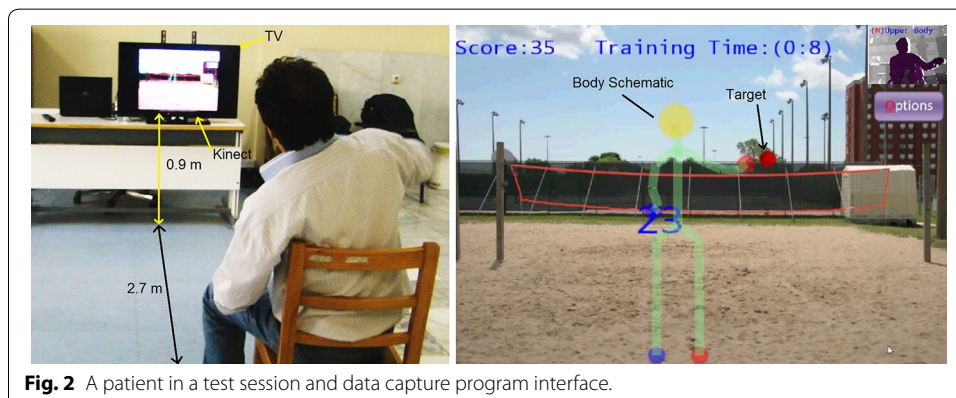
and Waist position) were recorded for further analysis using Kinect's skeleton tracking driver. A picture of one of the patients in a test session and the data capture program interface is shown in Fig. 2. Data capture program guided the patient through training exercises. In this figure a red ball is approaching the patient which implies that the patient should intercept the ball using his red (right) hand.

The protocol of reliability assessment

The intra-session and inter-session variabilities were adopted for this study, where the former concerns the same session and the latter deals with day-by-day reliability of the system.

Each test session was divided into 4 sub-sessions which lasted 5 min each with 2-min rest intervals. In the test period, subjects were told to catch as many balls as they could. The tests were held twice a week for the patients. They were asked to continue this routine for at least 4 weeks. Therefore, each patient had a minimum of 8 sessions and 32 sub-sessions of training. In order to lower the effects of any misunderstandings about the instructions and adaptation to the virtual environment of the tests, the first test session was considered as an orientation. Therefore, the intra-session reliability analysis was applied to performance indices from the second session.

The inter-session reliability approach was applied to the second sub-session of the last two sessions of each patient's tests. This choice was to lower the systematic error related to the patient's progress as its slope leveled off with time. It should be noted that in this study, all patients did their conventional physical therapy sessions two or three times a week and these tests took place parallel with their usual rehabilitation program.

Fig. 2 A patient in a test session and data capture program interface.

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 5 of 13

Performance indices measurement

The positions of upper body joint centers were recorded by Kinect at a sampling frequency of about 30 Hz during the tests. The sampling frequency of Kinect fluctuated between 25.61 and 34.72 Hz with a mean frequency of 29.9 Hz and a standard deviation of 2.67 Hz.

In order to extract performance indices, hand velocity, acceleration and jerk had to be calculated. In these calculations, Kinect's reference frame was used as the main coordinate system. Since numerical derivation to find velocity and acceleration intensifies the noise, it was essential to smoothen the data before any derivation. For this purpose, a B-spline (which is a piecewise polynomial function of order k [30]) was fitted to the position data. The order of B-spline used in this study was 6. The output of this method was a smooth function which was differentiable up to 5 times.

Subsequently, indices of movement performance for each hand, were extracted and calculated as listed below:

1. Mean velocity (MV): the mean value of the hand velocity [15] is defined as

$$MV = \frac{\sum_{i=1}^{N} V_i}{N} \tag{1}$$

   where V_i is the hand velocity at the ith sample of data, and N is the number of data samples.

2. Normalized mean speed (NMS): it is the mean value of the hand velocity divided by its maximum value [15],

$$NMS = \frac{MV}{V_{max}} \tag{2}$$

3. Normalized speed peaks (NSP): speed peaks are points where acceleration trajectory crosses the x-axis. NSP is defined as the number of speed peaks divided by the number of data samples [15],

$$NSP = \frac{\text{Number of speed peaks}}{N} \tag{3}$$

4. Logarithm of dimensionless jerk (LJ): it is the logarithm of median of hand's dimensionless jerk [19],

$$LJ = \text{Log} \left(\text{median} \left(\frac{(\int_{t_1}^{t_2} (\dddot{X}^2 + \dddot{Y}^2 + \dddot{Z}^2) dt)(t_2 - t_1)}{V_{mean}^2} \right) \right) \tag{4}$$

   where t_1 is the start time and t_2 is the end time of the movement, X, Y and Z are the positions of the hand measured by Kinect, and V_{mean} is the mean velocity in the movement.

5. Curvature (C): it is the logarithm of the median of hand's path curvature [17],

$$C = \text{Log} \left(\text{median} \left(\sqrt{\frac{(\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2)(\ddot{X}^2 + \ddot{Y}^2 + \ddot{Z}^2) - (\dot{X}\ddot{X} + \dot{Y}\ddot{Y} + \dot{Z}\ddot{Z})^2}{(\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2)^3}} \right) \right) \tag{5}$$

Mobini *et al. BioMed Eng OnLine (2015) 14:75*

Page 6 of 13

where *X, Y* and *Z* are positions of the hand measured by Kinect.

6. Spectral arc length (*SAL*): it is the negative arc length of the frequency-normalized Fourier magnitude spectrum of the speed profile [18],

$$SAL = -\int_0^{\omega_c} \sqrt{\left(\frac{1}{\omega_c}\right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega}\right)^2} d\omega, \hat{V}(\omega) = \frac{V(\omega)}{V(0)} x \tag{6}$$

where $V(\omega)$ is the Fourier magnitude spectrum of $V(t)$, and ω_c is the frequency band occupied by the given movement ($\omega_c = 40\pi \; rad/s$).

7. Shoulder angle with body (*SA*): the mean value of arm angle with body,

8. Elbow angle (*EA*): the mean value of elbow angle.

Both patients and healthy subjects did the exercises following the program guidance. All of the performance indices were measured in every reaching movement and their overall average in each sub-session was calculated so as to obtain one value for each training sub-session.

Statistical analysis

This study has followed the statistical methods used by Colombo et al. [31] who assessed performance indices reliability measured by robotic systems.

To reach a general view of the variability of the measured performance indices, a scatter plot was drawn for each of the indices measured over two sub-sessions of a single session and repeated measures ANOVA (analysis of variance) was used to calculate the reliability [32, 33].

Naturally, tests such as those carried out in this study involve a parameter of learning, and therefore, existence of learning-related error is taken for granted and not by any means any flaw of the tests. Nevertheless, to reduce the effects of this error on the reliability of intra-session indices, the first sub-session of each test was overlooked to let the patients adapt to the circumstances of the test, and the last 3 sub-sessions were investigated. As for the inter-session reliability of the tests, the last two sessions were analyzed to lower the influence of the learning procedure.

To model the system, *N* subjects with *M* repeated measurements of continuous variable *P* were considered. A mathematical model for measurements of *P* was considered here as:

$$P_{ij} = T + t_i + S_{ij} + R_{ij} \tag{7}$$

where P_{ij} is the *j*th measurement ($j = 1, \ldots, M$) made on the *i*th subject ($i = 1, \ldots, N$), *T* is the true value of the variable, t_i is the subject's effects on the true value, S_{ij} is the systematic error and R_{ij} is the random error. t_i, S_{ij}, and R_{ij} are independent random errors which are normally distributed with means of 0 and variances of σ_t^2, σ_S^2, and σ_R^2, respectively [31, 34, 35]. The reliability of parameter *P* can be calculated using intra-class correlation coefficient (*ICC*) defined as [35]:

$$\text{Reliability} = \frac{\text{Between subjects variability}}{\text{Between subjects variability} + \text{error}} \tag{8}$$

$$R_u = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_S^2 + \sigma_R^2} \tag{9}$$

Since this study did not consider the systematic error, Eq. (9) was reduced to Eq. (10):

$$R = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_R^2} \tag{10}$$

Both inter-session and intra-session reliabilities were calculated as positive values between 0 and 1. A model of repeated-measures ANOVA was used to determine this parameter. In this method σ_t and σ_r were determined as:

$$\sigma_t = \frac{MS_S - MS_E}{k} \tag{11}$$

$$\sigma_R = MS_E \tag{12}$$

where MS_S is the subjects difference mean square and calculated based on differences among subjects in measurements of each trial, MS_E is error mean square which was calculated based on the difference between evaluations of one subject's trials and k is the number of trials which is 2 for this study.

By replacing Eqs. (11) and (12) in Eq. (10), McGraw and Wong's [33] 2-way fixed model equation (C, 1) was obtained as below:

$$R = \frac{MS_S - MS_E}{MS_S + (k-1)MS_E} \tag{13}$$

Moreover, the standard error of measurement (*SEM*) was calculated which makes an absolute index of reliability available, and allows for the quantification of each measurement's precision. *SEM* was defined as the square root of the mean square from ANOVA results. *SEM* has the same units of each measured indices and encompasses components of random and systematic error of measurement.
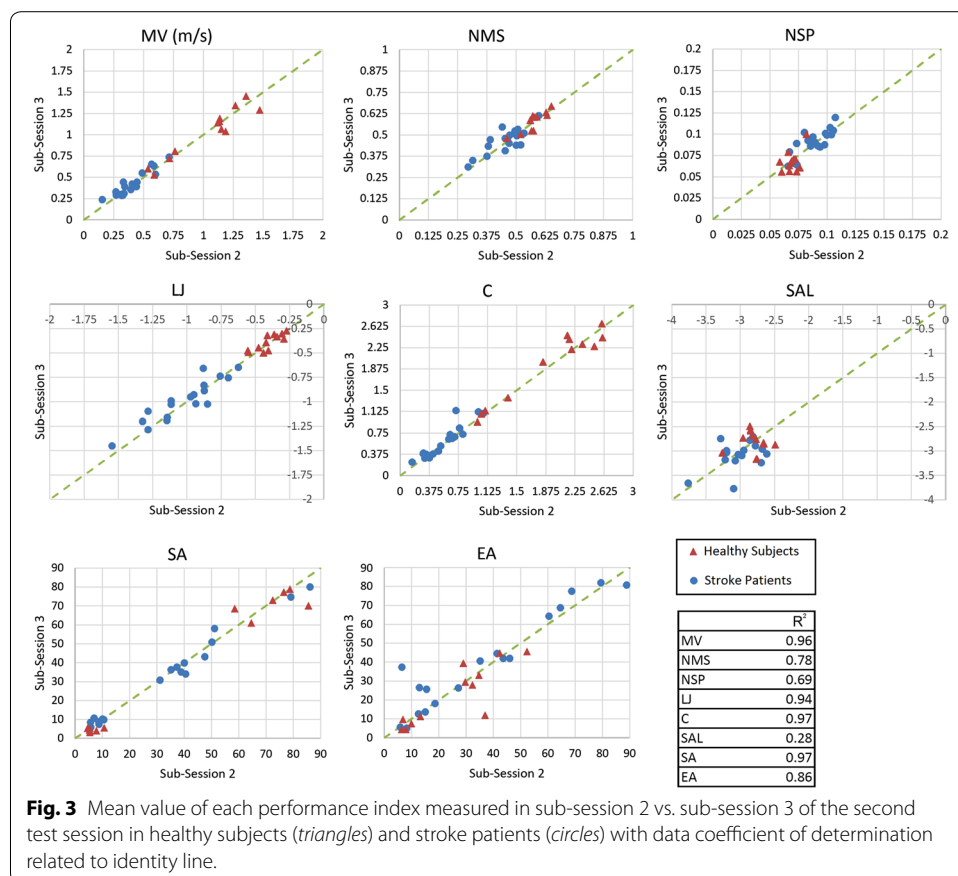
$$SEM = \sqrt{MS_E} \tag{14}$$

Furthermore, *SEM*'s coefficient of variation (*CV*) was calculated in order to optimize the result comparison in cases of various unites and scales [36]. The *CV* of *SEM* was defined as the ratio between *SEM* and each index's overall mean and was presented as a percentage. Another parameter, calculated in this study, was the minimal detectable difference (*MDD* [33]) which indicates the minimum difference required to state a significant change in an index. Equation (15) demonstrates how *MDD* was calculated [33]. This parameter enables the examiner to realize whether any noticeable change has occurred in the movement quality index.

$$MDD = SEM \times 1.96 \times \sqrt{2} \tag{15}$$

Results

As described in the last section, to obtain an overview of the results, the scatter plot of each performance index was plotted for two sub-sessions of the second test session as shown in Fig. 3. Each part of this figure reports the mean value of one index measured in sub-session 2 vs. sub-session 3 for all patients and healthy subjects. Since the indices were measured in two consecutive sub-sessions, consistency of the index required
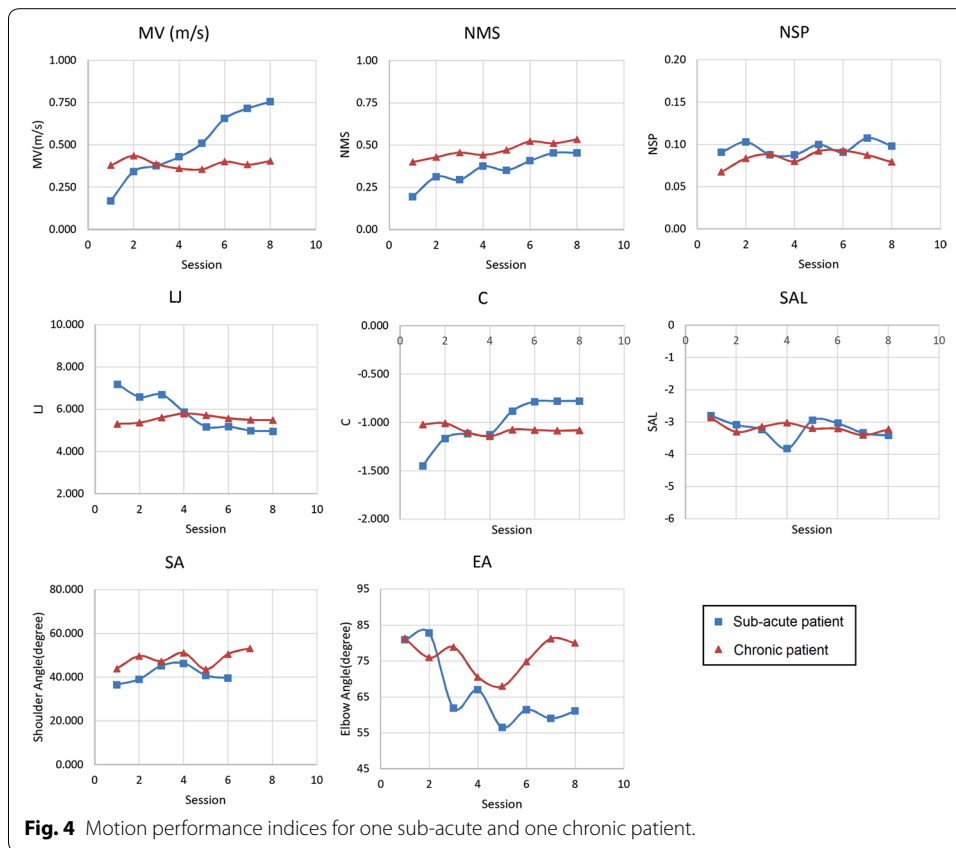
Mobini *et al. BioMed Eng OnLine (2015) 14:75*

Page 8 of 13

Fig. 3 Mean value of each performance index measured in sub-session 2 vs. sub-session 3 of the second test session in healthy subjects (*triangles*) and stroke patients (*circles*) with data coefficient of determination related to identity line.

the data points to be located close to the identity line. To compare performance indices, coefficient of determination (R^2) [37] was found with respect to the identity line and shown in Fig. 3. The figure shows that, except *SAL*, the other indices are very close to the identity line, implying better consistencies. Among these indices, *MV*, *LJ*, *C* and *SA* show the best performance with R^2 greater than 0.9. This figure also demonstrates a good separation between healthy subjects and patients in *MV*, *NMS*, *NSP*, *LJ* and *C*. This means that these indices are better nominees for assessing the patient's state and progress. Also, the values of *MV*, *LJ* and *C* are separated more widely along the identity line which may imply a better resolution in distinguishing between the patients' status.

Motion performance indices for one sub-acute and one chronic patient in 8 sessions are shown in Fig. 4. In this figure, the mean value of each performance index in the four sub-sessions is considered as the value for that session. As shown in this figure, for the sub-acute patient *MV*, *NMS*, *LJ* and *C* change uniformly in the period of the test which indicates patient's progress in this span of time. On the other hand, for the chronic patient, the variation of the indices is not significant which is because of his lower rate of recovery relative to the sub-acute patient.

The mean value, standard deviation, and intra-session test–retest reliability of the results measured by Kinect from healthy subjects and stroke patients are presented in Tables 1 and 2. The intra-session test–retest reliability of the results is expressed by *ICC*, *SEM*, *CV* and *MDD* as defined in the last section. These tables show high *ICC* values

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 9 of 13

Fig. 4 Motion performance indices for one sub-acute and one chronic patient.

Table 1 Intra-session reliability parameters in healthy subjects

Performance Indices	Mean ± SD	ICC	SEM	CV	MDD
MV (m/s)	1.033 ± 0.296	0.96	0.083	8.02	0.230
NMS	0.576 ± 0.053	0.86	0.024	4.19	0.067
NSP	0.068 ± 0.009	0.42	0.006	9.35	0.018
LJ	−0.399 ± 0.087	0.81	0.034	8.44	0.093
C (m)	1.923 ± 0.609	0.97	0.114	5.94	0.316
SAL	−2.812 ± 0.202	0.34	0.116	4.12	0.322
SA (°)	38.86 ± 34.45	0.98	2.216	5.70	6.144
EA (°)	23.75 ± 15.51	0.86	4.371	18.4	12.11

Table 2 Intra-session reliability parameters in stroke patients

Performance indices	Mean ± SD	ICC	SEM	CV	MDD
MV (m/s)	0.415 ± 0.141	0.93	0.057	13.8	0.159
NMS	0.459 ± 0.075	0.81	0.045	9.77	0.124
NSP	0.090 ± 0.014	0.77	0.007	8.33	0.021
LJ	−1.001 ± 0.228	0.91	0.087	8.68	0.242
C (m)	0.588 ± 0.269	0.91	0.07	11.9	0.194
SAL	−3.136 ± 0.385	0.52	0.31	9.85	0.856
SA (°)	32.69 ± 23.64	0.99	4.14	12.7	11.53
EA (°)	37.86 ± 25.92	0.94	5.84	15.4	16.19

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 10 of 13

(more than 0.8) for all indices except *NSP* and *SAL* for both patients and healthy subjects. The *CV* in healthy subjects is mostly lower than stroke patients which points to the higher consistency of healthy subjects' movements. The worst performance of *SEM* is for *EA* which has more than 15 % of *CV* for both patients and healthy subjects. A comparison of the *MDD* and the difference between mean values of indices for healthy subjects and stroke patients indicates that there is a meaningful difference between these numbers. This comparison shows that one can distinguish patients from healthy subjects using most of these indices.

The mean value, standard deviation and inter-session test–retest reliability of the results measured by Kinect from stroke patients are presented in Table 3. This table, also, shows high *ICC* values (more than 0.9) for all indices except *NSP*, *NMS* and *SAL*. In this table, similar to the others, the worst performance of *SEM* belongs to *EA* which has more than 16 % of *CV*. Comparing mean values of indices in this table and Tables 1 and 2 shows that mean values of indices from Table 3 lie between those from Tables 1 and 2. For a better comparison, the mean value and standard deviation of the differences between indices from the second session and the last session of therapy are shown in Table 4. As reported in Table 4, only in *MV, LJ* and *C*, the mean difference is more than *MDD*. This shows that only these three indices present a meaningful variation in 1 month of therapy. Thus using these indices, one is able to measure patients' progress in this period.

Table 3 Inter-session reliability parameters in stroke patients

Performance indices	Mean ± SD	ICC	SEM	CV	MDD
MV (m/s)	0.669 ± 0.336	0.94	0.086	12.8	0.239
NMS	0.459 ± 0.074	0.6	0.034	7.36	0.093
NSP	0.085 ± 0.012	0.71	0.007	7.9	0.019
LJ	−0.779 ± 0.293	0.95	0.089	11.4	0.247
C (m)	0.951 ± 0.517	0.96	0.141	14.7	0.389
SAL	−3.19 ± 0.996	0.12	0.413	12.9	1.144
SA (°)	30.32 ± 24.46	0.96	4.22	13.9	11.71
EA (°)	33.12 ± 19.33	0.92	5.49	16.58	15.22

Table 4 Change of indices in stroke patients after 1 month of rehabilitation

Performance indices	Change (mean ± SD)	MDD
MV (m/s)	0.292 ± 0.303	0.239
NMS	0.001 ± 0.126	0.093
NSP	−0.005 ± 0.021	0.019
LJ	0.285 ± 0.222	0.247
C (m)	0.443 ± 0.399	0.389
SAL	−0.058 ± 1.061	1.144
SA (°)	−8.52 ± 18.19	11.71
EA (°)	−8.71 ± 15.61	15.22

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 11 of 13

Discussions

The results, as summarized in Tables 1, 2, 3, 4, indicate an acceptable intra-session and inter-session reliability for measuring *MV, LJ, C, SA* and *EA*. They are comparable to those computed with robotic systems [31] or clinical assessment scales reported in the literature [38, 39]. The measured *ICC* of Kinect and robotic devices reported by Colombo et al. [31] are summarized in Table 5. As shown in the table, the *ICC* values of *MV* are quite close. Although, the *ICC* of *LJ, C, SA* and *EA* are not reported in the Colombo's work, they have values close to the *ICC* of *MV* when measured by Kinect. The *ICC* of the other indices, i.e. *NMS, NSP* and *SAL*, are relatively lower when measured by Kinect and hence not as reliable as the other indices.

Comparing the results of Tables 1 and 2 shows that there are significant differences between the mean values of all indices for healthy subjects and stroke patients. Also, all of the differences are larger than *MDD*, pointing to the fact that all of these indices while measured by Kinect can be used in separating patients from healthy subjects. However, comparison between the results of Tables 2 and 3 which is summarized in Table 4 implies that only the variations of *MV, LJ* and *C* are significant after 1 month of therapy and the other indices are not capable of reflecting patients' progress in this period of time.

A comparison of the results of Tables 3 and 1 shows that even after 1 month of therapy, there is a significant difference between all mean values of the indices for healthy subjects and stroke patients. This means that even after 1 month of therapy, stroke patients and healthy subjects can be easily separated by all of the indices although none of *NMS, NSP, SAL, SA* and *EA* are sufficiently sensitive for assessing patients' progress in this period of rehabilitation. Also, this indicates that saturation doesn't happen when using *MV, LJ* and *C* as patients' progress indicators. After 1 month of therapy, they still maintain a considerable distance from a healthy subject.

As stated above, *NMS, NSP, SAL, SA* and *EA* seem to be incapable of reflecting the patient's progress when measured by Kinect. This is in conflict with a number of other related studies [15, 18, 31] that show significant changes in these indices during 1–6 months of rehabilitation. This has to be due to Kinect's lower accuracy in measuring

Table 5 Measured *ICC* of Kinect and Colombo et al. [31] results for robotic devices

Performance indices	Healthy *ICC*	Patients *ICC*		Healthy *ICC* [31]	Patients *ICC* [31]	
	I1[a]	I1	I2[b]	I1	I1	I2
MV (m/s)	0.96	0.93	0.94	0.95	0.97	0.93
NMS	0.86	0.81	0.6	0.99	0.94	0.91
NSP	0.42	0.77	0.71	0.85	0.82	0.95
LJ	0.81	0.91	0.95	–	–	–
C (m)	0.97	0.91	0.96	–	–	–
SAL	0.34	0.52	0.12	0.92	0.83	0.95
SA (°)	0.98	0.99	0.96	–	–	–
EA (°)	0.86	0.94	0.92	–	–	–

[a] Intra session reliability.

[b] Inter session reliability.

Mobini *et al. BioMed Eng OnLine (2015) 14:75*

Page 12 of 13

those particular indices which could lower their sensitivity. In related literatures, clinical measurement systems were used which are more accurate than Kinect [25, 40–42].

The inter-session and intra-session test–retest reliability of indices measured by Kinect for healthy subjects had lower than 4% difference. Due to this small difference, the inter-session test–retest reliability results are not reported for the sake of brevity.

As Fig. 4 demonstrates, the learning effect on most of the performance indices of subacute patients are significant in the first month of therapy. In later studies, it is suggested to prolong the test time to 6 months in order to lower the effects of patient's learning on inter-session reliability.

Finally it should be remembered that the objective of this study was to investigate the reliability of performance indices when measured by Kinect. In future works, it is necessary to study the measurement accuracy of these indices in comparison to popular clinical devices such as marker-based optical and inertial ones. Furthermore, determination of the correlations between the performance indices and the clinical measures of the subject's quality of motion is another interesting subject which is being pursued by the authors in a separate study.

Conclusions

In this study, a systematic approach was followed to assess the intra-session and inter-session reliability of performance indices measured by Microsoft Kinect. The results of this study showed that, among the performance indices, *MV*, *LJ*, *C*, *SA* and *EA* had more than 0.9 *ICC* together with an acceptable *SEM* and *CV* in both stroke patients and healthy subjects across sessions. However, only *MV*, *LJ* and *C* showed significant variations after 1 month of therapy and hence concluded to be suitable for progress evaluation of patients. The results are promising for the development of home-based rehabilitation systems using Kinect as an affordable motion capture sensor.

References
1. Tsoupikova D, Stoykov N, Kamper D, Vick R. Virtual reality environment assisting post stroke hand rehabilitation: case report. Stud Health Technol Inform. 2012;184:458–64.
2. Wade E, Winstein CJ. Virtual reality and robotics for stroke rehabilitation: where do we go from here? Topics Stroke Rehabil. 2011;18(6):685–700.
3. Deutsch JE, Borbely M, Filler J, Huhn K, Guarrera-Bowlby P. Use of a low-cost, commercially available gaming console (Wii) for rehabilitation of an adolescent with cerebral palsy. Phys Ther. 2008;88(10):1196–207. doi:10.2522/ptj.20080062.
4. Meldrum D, Glennon A, Herdman S, Murray D, McConn-Walsh R. Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii® Fit Plus. Disabil Rehabil Assist Technol. 2012;7(3):205–10.

Mobini *et al. BioMed Eng OnLine* (2015) 14:75

Page 13 of 13

5.   Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil. 2012;9(12):1–17.

6.   Bao X, Mao Y, Lin Q, Qiu Y, Chen S, Li L, et al. Mechanism of Kinect-based virtual reality training for motor functional recovery of upper limbs after subacute stroke. Neural Regen Res. 2013;8(31):2904–13.

7.   Microsoft Kinect. http://www.xbox.com/en-US/kinect. Accessed Feb 2015.

8.   Nintendo Wii and balance Board., http://www.nintendo.com/wii. Accessed Feb 2015.

9.   Vicon. http://www.vicon.com/. Accessed Feb 2015.

10.   Optotrak. http://www.ndigital.com/msci/products/optotrak-certus/, http://www.vicon.com/. Accessed June 2015.

11.   Wii Fit Plus. http://wiifit.com/. Accessed Feb 2015.

12.   SeeMe System. http://www.virtual-reality-rehabilitation.com/products/seeme/what-is-seeme. Accessed Feb 2015.

13.   VirtualRehab System., http://www.virtual-reality-rehabilitation.com/, http://virtualrehab.info/product/#innovative-technology. Accessed Feb 2015.

14.   JINTRONIX System. http://www.jintronix.com/. Accessed Feb 2015.

15.   Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, et al. Movement smoothness changes during stroke recovery. J Neurosci. 2002;22(18):8297–304.

16.   van Dokkum L, Hauret I, Mottet D, Froger J, Métrot J, Laffont I. The contribution of kinematics in the assessment of upper limb motor recovery early after stroke. Neurorehabil Neural Repair. 2014;28(1):4–12. doi:10.1177/1545968313498514.

17.   Osu R, Ota K, Fujiwara T, Otaka Y, Kawato M, Liu M. Quantifying the quality of hand movement in stroke patients through three-dimensional curvature. J NeuroEng Rehabil. 2011;8(1):1–14. doi:10.1186/1743-0003-8-62.

18.   Balasubramanian S, Melendez-Calderon A, Burdet E. A robust and sensitive metric for quantifying movement smoothness. IEEE Trans Biomed Eng. 2012;59(8):2126–36.

19.   Hogan N, Sternad D. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. J Mot Behav. 2009;41(6):529–34.

20.   Xsens. http://www.xsens.com/. Accessed Feb 2015.

21.   Wikstrom EA. Validity and reliability of Nintendo Wii Fit balance scores. J Athl Train. 2012;47(3):306.

22.   Park D-S, Lee G. Validity and reliability of balance assessment software using the Nintendo Wii balance board: usability and validation. J Neuroeng Rehabil. 2014;11(1):99.

23.   Kim J, Nam KW, Jang IG, Yang HK, Kim KG, Hwang J-M. Nintendo Wii remote controllers for head posture measurement: accuracy, validity, and reliability of the infrared optical head tracker. Invest Ophthalmol Vis Sci. 2012;53(3):1388–96.

24.   Obdrzalek S, Kurillo G, Ofli F, Bajcsy R, Seto E, Jimison H et al., editors. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. 2012.

25.   Mobini A, Behzadipour S. Saadat Foumani M. Accuracy of Kinect's skeleton tracking for upper body rehabilitation applications. Disabil Rehabil Assis Technol. 2014;9(4):344–52. doi:10.3109/17483107.2013.805825.

26.   Bonnechère B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Moiseev F et al. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait and Posture. 2014;39(1):593–98. doi:10.1016/j.gaitpost.2013.09.018.

27.   Clark RA, Pua Y-H, Fortin K, Ritchie C, Webster KE, Denehy L, et al. Validity of the Microsoft Kinect for assessment of postural control. Gait Posture. 2012;36(3):372–7.

28.   Huber ME, Seitz AL, Leeser M, Sternad D. Validity and reliability of Kinect skeleton for measuring shoulder joint angles: a feasibility study. Physiotherapy. 2015;. doi:10.1016/j.physio.2015.02.002.

29.   Elgendi M, Picon F, Magnenat-Thalmann N, Abbott D. Arm movement speed assessment via a Kinect camera: a preliminary study in healthy subjects. Biomed Eng Online. 2014;13(1):88.

30.   De Boor C. A practical guide to splines. New York: Springer-Verlag; 1978.

31.   Colombo R, Cusmano I, Sterpi I, Mazzone A, Delconte C, Pisano F. Test-retest reliability of robotic assessment measures for the evaluation of upper limb recovery. Neural Syst Rehabil Eng IEEE Trans. 2014;. doi:10.1109/TNSRE.2014.2306571.

32.   Atkinson G, Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217–38. doi:10.2165/00007256-199826040-00002.

33.   Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the *SEM*. J Strength Cond Res. 2005;19(1):231–40.

34.   Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Stat Med. 2002;21(22):3431–46.

35.   Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420.

36.   Hopkins W. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15. doi:10.2165/00007256-200030010-00001.

37.   Steel RGD, Torrie JH. Principles and procedures of statistics: with special reference to the biological sciences. 1960.

38.   Gladstone DJ, Danells CJ, Black SE. The Fugl–Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16(3):232–40.

39.   Platz T, Pinkowski C, van Wijck F, Kim I-H, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil. 2005;19(4):404–11.

40.   Dutta T. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. Appl Ergon. 2012;43(4):645–9. doi:10.1016/j.apergo.2011.09.011.

41.   Fern'ndez-Baena A, Susín A, Lligadas X, editors. Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on. 2012; IEEE.

42.   Webster D, Celik O, editors. Experimental evaluation of Microsoft Kinect's accuracy and capture rate for stroke rehabilitation applications. Haptics Symposium (HAPTICS), 2014 IEEE; 23–26 Feb 2014.