

A preliminary characterization of physical therapist visual behaviors during standing balance tasks using eye tracking

Emma Nigrelli

Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI, USA
enigrell@umich.edu

Leia Stirling

Department of Industrial and
Operations Engineering
Robotics Department
University of Michigan
Ann Arbor, MI, USA
leias@umich.edu

Kathleen Sienko

Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI, USA
sienko@umich.edu

Abstract—Visual observations provide an important foundation for physical therapist (PT) decision making in balance rehabilitation. This study used eye tracking to identify areas of the body focused on by PTs while evaluating standing balance across different exercise difficulties. Eye tracking data were obtained from five PTs who observed and rated older adults performing standing balance exercises, using a five-point scale. Excluding ratings of five, PTs had an average of approximately 215% more mean visits and an average of approximately 195% longer mean average visit durations to the lower extremities (LE) compared to the other regions of the body. However, visits to the LE shortened with increasing balance challenge, with approximately 75% shorter mean average visit durations to the LE for exercises rated as five compared to exercises rated as one. The number of visits to the head increased with balance challenge, with approximately 230% more mean number of visits to the head for ratings of five compared to ratings of one. Shortened visit durations and an increased number of visits to the upper body suggest an increase in whole body scan patterns with increasing challenge level.

Keywords—eye tracking, physical therapist, clinical decision making, balance rehabilitation

I. INTRODUCTION

Physical therapists (PTs) play an essential role in balance rehabilitation, including identifying impairments in balance and their causes, such as those that may impact safety or daily activities [1], as well as making decisions for future treatment. Existing frameworks for PT decision-making are built upon both prior knowledge and observations or predictions of patient performance, where clinicians use this information to develop and continually update hypotheses for patient weaknesses [2], [3], [4]. While evaluating balance, PTs consider the presence or absence of various compensatory reactions, such as ankle, knee, hip, and stepping strategies, which are responses to increasing exercise difficulty [5]. These observations and other insights into the PT decision-making process can be captured through their documented evaluations and reports. In addition to clinical functional assessment scores or ratings, eye tracking provides further insight into PTs' specific observational practices.

Eye tracking measures the location of foveal vision, assumed to be the focus of attention [6]. Previous studies have used eye tracking to investigate clinicians' thought processes and performance across a variety of clinical settings. Applications include understanding how clinicians read ECGs [7], [8], make diagnoses in pediatric neurology [9], and evaluate chest x-rays [10]. Balslev et al. [9] determined expert clinicians making pediatric neurologic diagnoses focus longer on relevant areas to verify their hypotheses, while Desvergez et al. [11] concluded the high number of fixations seen in expert clinicians indicates frequent iteration on mental models during an emergency.

PT visual behavior has also been investigated, particularly how scan patterns vary with experience level. Liu et al. [12] found experienced PTs exhibit "rhythmical" visual behavior, looking between specific areas of the body during gait. McDuff et al. [13] and Hayashi et al. [14] compared PT visual behavior across experience level as the PTs viewed videos of a neurologic patient performing sit-to-stand and abnormal gait, respectively. Both studies found experienced PTs had more and shorter fixations on relevant areas compared to novices. Although Hayashi et al. [14] compared PT visual behavior with the gait features they identified, previous studies have focused on visual behavior relative to PT experience level, rather than visual behavior relative to the PT decisions and evaluations made.

Understanding the relevant areas and scanning behaviors used in relation to patient performance will build upon the existing PT decision-making frameworks by describing how and what information is gathered during observations and whether these visual strategies change with exercise difficulty. As such, the goal of this study was to identify the areas of the body PTs are most attentive to while evaluating standing balance in older adults, relative to their provided ratings of balance performance.

II. METHODS

Clinician and healthy older adult pairs ($n=5$ pairs) were recruited for this study, similar to sample sizes in related studies [10], [12]. The recruited PTs (5 female) had an average of 11.8 (SD: 9.7) years of experience and an average of 8.4 (SD: 3.9) years of balance or neurologic disorders experience. The

recruited balance participants (2 male, 3 female) had a mean age of 73.8 (SD: 3.4) years. The University of Michigan Institutional Review Board reviewed and approved the study protocol (HUM00219572) and written informed consent was provided by all study participants.

The PTs viewed the exercise participants performing standing balance tasks. PTs were provided an initial location approximately three meters from the balance participant with approximately a 45 degree angle between the two participants; however, the PTs were permitted to move freely while watching the exercise. A study team member acted as a spotter. PTs were told the balance participants were healthy older adults and that they would be asked provide ratings of performance and their rationales for these ratings. Ratings were given on the 1-5 scale used by Bao et al. [15], where one denoted “independent with no sway” and five denoted “unable to maintain position with assist or step out” during the first half of the exercise.

Balance participants performed three trials of between four and eight standing balance exercises. Exercise difficulty varied by surface (firm, compliant), head movement (none, pitch, yaw), visual input (eyes open, eyes closed), and stance (feet apart, feet together, semi-tandem, tandem, single leg) [16], similar to the protocol described in Ferris et al. [17]. Each balance participant performed a unique combination of exercises; however, some exercises were repeated across participants. The exercise duration was defined between the balance participant’s start and stop cues, with a mean trial length of 30.8 (SD: 0.54) seconds.

To track visual behavior, the PTs wore eye tracking glasses (Tobii Pro Glasses 2, Tobii, Danderyd, Sweden), which were calibrated with Tobii’s single point calibration card held approximately one meter from the participant. Calibrations were repeated approximately halfway through the testing session. Gaze data were collected for 26 sets of three exercise trials across all PTs, with two PTs viewing eight exercise sets, two PTs viewing four exercise sets, and one PT viewing two sets of exercises (due to a recording being corrupted). Seven areas of interest (AOIs) were defined (Fig. 1). For single leg trials, a non-supporting foot AOI was added, while the “feet” and “knees” AOIs contained only the supporting foot and knee, respectively. The AOIs contained the largest possible portion of the segment, but did not overlap. AOIs were defined by two different team members with one team member reviewing all completed AOIs.

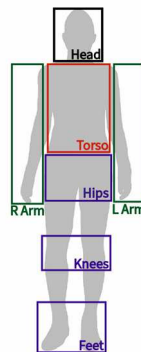


Fig. 1. Balance participant AOI placement.

Metrics were calculated using Tobii Pro Lab Analyzer software (Tobii, Danderyd, Sweden). The data were processed using Tobii’s built-in I-VT (Attention) filter, including a 100°/s velocity threshold, a 60ms minimum fixation duration, and a maximum separation between fixations of 75ms and 0.5° [18]. The filter used 3 sample moving median window for noise reduction and the data were interpolated for a maximum gap length of 75ms. The average visit duration and the number of visits were calculated for each AOI for each trial [6]. Visits may include multiple fixations and were defined from the first to last fixation in a region [6]. For analysis, the feet, non-supporting foot, knees, and hips AOIs were combined to create the “lower extremities” region (LE), while the left and right arm AOIs were combined into the “upper extremities” region (UE). Visits were also considered to the head and torso regions. Due to small sample sizes, significance was not assessed; rather, effect sizes were calculated via Cohen’s d, with a large effect being greater than 0.80 and a medium effect being greater than 0.50 [19].

III. RESULTS

Across the 78 exercises viewed by the PTs, 38 were rated as one, 26 were rated as two, eight were rated as three, four were rated as four, and two were rated as five. Gaze data were obtained for at least one eye for a minimum of 70% (mean: 92.8%) of the trial duration.

For ratings of one through three, the LE received the highest number of visits, followed by the torso, the head, and the UE (Fig. 2). Ratings of one and two had the longest visits to the LE, followed by the head and torso with similar visit durations. However, ratings of three had similar visit durations between the head, LE, and torso, with visits to the LE being shorter than those for both ratings of one and two. The UE had the shortest visits for ratings one through three. Compared to all other ratings, the ratings of one had the longest visits to the LE. Table 1 shows the absolute difference of the means between all combinations of AOIs for the number of visits and average visit duration by each rating and the corresponding effect size.

Similar to ratings of one through three, ratings of four had the fewest visits to the UE and the most visits to the LE. However, the head and the torso had a similar number of visits, with the head receiving more visits than ratings of one through three. The LE had the longest visits for ratings of four, followed by the head, torso, and UE. Ratings of five had a high number of visits to both the LE and the Head, however, the head received the most visits, unlike the lower ratings that visited the LE most. Five ratings had a low number of visits to the Torso and the UE, with the torso receiving fewer visits and the UE receiving more visits than ratings one through four. Unlike all of the other rating levels, ratings of five had the longest visits to the head, followed by the LE. The Torso and UE had the shortest and similar visit durations. Across all AOIs, ratings of five had the shortest average visit duration, especially for the Torso and the LE, with the head having a similar duration to ratings of one and the UE having a similar visit duration to ratings of two and four.

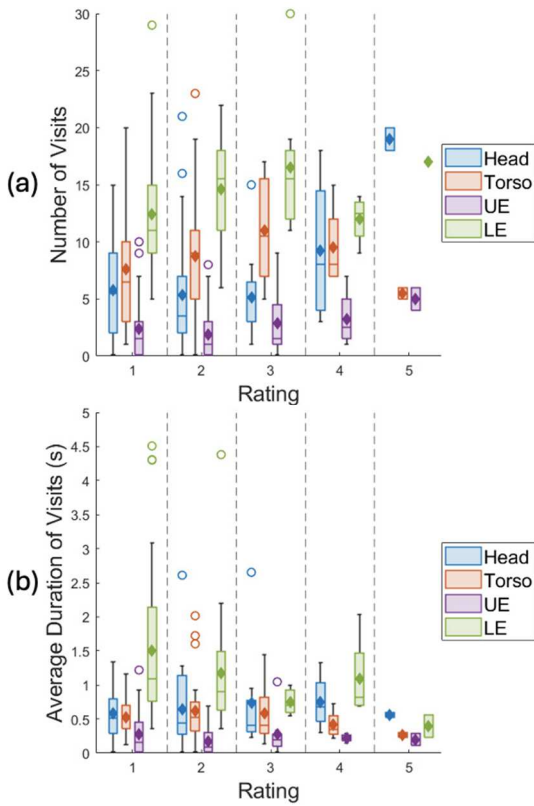


Fig. 2. Plot of number of visits across all exercise trials by rating (a) and plot of average visit duration across all exercise trials by rating (b). Boxplots show the interquartile range, median, and range, while the diamonds represent the mean.

IV. DISCUSSION

Overall, the LE had the most visits across all ratings, with the exception of ratings of five, which had more visits to the head. This prominence of the LE indicates the LE consistently provides relevant and informative context for patients' performance, following the assumption that the number of visits is indicative of attention levels [20]. This high frequency of visits to the LE is consistent with the many lower body balance strategies that may be employed, including hip, knee, ankle, and stepping strategies [1], [5]. Conversely, the UE had low informativeness, with the fewest and shortest visits across all ratings, except five which had a similar number of visits between the UE and torso. The low UE relevance may be attributed to the limited number of difficult exercises performed, as indicated by

few ratings of five, because arm strategies are often used to prevent falls [21].

Ratings of one had the longest visits to the LE across all ratings, indicating sustained visits to the lower body to gather information for the easiest exercises. Similarly, ratings of two had the most and longest visits to the LE, accompanied by fewer and shorter visits to the head and torso. This pattern suggests PTs observed all areas of the body for exercises rated as a two; however, they returned to the LE more frequently and for longer due to the presence of relevant information. The LE prioritization indicates the salience of lower body strategies for the easiest exercises, consistent with previous work finding knee and ankle strategies are more often utilized and successful at lower challenge levels [5].

Despite small sample sizes, observations can still be made for ratings of three through five to inform hypotheses for future studies with larger samples. Although ratings of three visited the LE the most, the visual behavior indicates increasing torso relevance, with ratings of three having the most visits to the torso compared to the other rating levels. Increased torso interest may be due to the increased torso sway associated with more challenging balance exercises [5]. Ratings of three had a similar duration of visits between the head, torso, and LE, further suggesting a growing importance of the upper body, as the PTs gave similar amounts of attention when looking at the three regions, despite having the most revisits to the LE. Although the LE still received the most visits for ratings of four, ratings of four had the fewest visits to the LE compared to all other ratings (although similar to ratings of one) and more visits to the head compared to ratings of one through three, again indicating a growing consideration of the whole body. Ratings of four had the longest visits to the LE compared to the other body regions, despite only having a few more visits to the LE compared to the torso and head. This gaze pattern could indicate whole-body scanning that slows in the LE to look for relevant strategies.

Finally, ratings of five had the most consideration of the whole body. The relatively short visits compared to the other ratings, as well as the grouping of the head and LE with the most visits and the torso and UE with the fewest visits, are indicative of a rapid, whole-body scanning pattern with many brief revisits to regions, especially the head and LE. This speed may contribute to the few visits to the torso and UE, as the eye movement may have been too rapid to be classified as a visit. Additionally, ratings of five had the most visits to the head compared to all other ratings, which may indicate higher relevance of facial expressions with increased challenge or the PTs verifying the patient's adherence to exercise components

TABLE I. MAGNITUDE OF THE DIFFERENCE OF MEANS BETWEEN EACH AOI COMBINATION FOR BOTH NUMBER OF VISITS AND AVERAGE DURATION OF VISITS, BY RATING. SHADING REPRESENTS EFFECT SIZE WITH NO EFFECT AS WHITE ($D < 0.5$), MEDIUM EFFECT ($0.5 < D < 0.8$) AS LIGHT GRAY AND LARGE EFFECT AS DARK GRAY ($D > 0.8$).

	Number of Visits					Avg. Duration of Visit (s)				
	1	2	3	4	5	1	2	3	4	5
Head-Torso	1.8	3.4	5.9	0.3	13.5	0.06	0.02	0.15	0.34	0.30
Head-UE	3.4	3.5	2.3	6.0	14.0	0.30	0.47	0.46	0.53	0.36
Head-LE	6.7	9.3	11.4	2.8	2.0	0.92	0.53	0.01	0.34	0.16
Torso-UE	5.2	6.9	8.1	6.3	0.5	0.24	0.45	0.31	0.19	0.06
Torso-LE	4.8	5.9	5.5	2.5	11.5	0.98	0.55	0.17	0.68	0.13
LE-UE	10.1	12.8	13.6	8.8	12.0	1.23	1.00	0.48	0.87	0.20

(e.g. maintaining eyes open or head movement), as such tasks are more challenging [16]. Across ratings three through five, the growing consideration for the upper body as exercise challenge increases is consistent with studies noting the general increase in whole-body balance reactions with increasing challenge levels [5]. Furthermore, the increase in scanning behaviors and rechecking regions as exercises become more difficult aligns with PTs confirming and updating their understanding, especially as more challenging exercises can be expected to have more unexpected or evolving balance reactions over the course of the trial. The results of this preliminary study may have implications for how novice PT and/or trainees are instructed.

Limitations of the methods included an uneven distribution of ratings across the different exercises (e.g., both ratings of five came from the same PT) and missing data from each trial. Although some of this data loss can be attributed to blinking, the PTs sometimes glanced over the top of the eye-tracking glasses, which may have underestimated visits to the head and torso. Additionally, the LE was defined as an aggregated AOI of the knees, hips, and feet, with space between each of these regions, which may have overestimated the number of visits to the LE if PTs' gaze fixated on the regions between the defined AOIs.

V. CONCLUSION

This study aimed to characterize the areas of the body that physical therapists are most attentive to while evaluating standing balance with respect to their ratings of balance performance. The results of this study showed that PTs had more visits to the LE, but the importance of the upper body increased with exercise difficulty. Additionally, as exercise difficulty increased, PTs exhibited increased scanning behavior to more frequently update their understanding, as indicated by shorter visits to the LE for higher rated exercises. Future research includes performing a study with a larger number of physical therapists, balance participants, and exercise ratings of all challenge levels. Additionally, future research should consider metrics such as scan path order to better understand PT strategies in detail, such as to confirm the presence of scanning behavior or revisits to regions.

ACKNOWLEDGEMENT

The authors acknowledge Danny Shin and Safa Jabri for contributions to data collection, Christina Sylvester for assistance with data processing, and Wendy Carender for providing physical therapist expertise.

REFERENCES

- [1] F. B. Horak, "Clinical assessment of balance disorders," *Gait Posture*, vol. 6, no. 1, pp. 76–84, Aug. 1997, doi: 10.1016/S0966-6362(97)00018-0.
- [2] M. Schenkman, J. E. Deutsch, and K. M. Gill-Body, "An Integrated Framework for Decision Making in Neurologic Physical Therapist Practice," *Phys. Ther.*, vol. 86, no. 12, pp. 1681–1702, Dec. 2006, doi: 10.2522/ptj.20050260.
- [3] J. M. Rothstein, J. L. Echternach, and D. L. Riddle, "The Hypothesis-Oriented Algorithm for Clinicians II (HOAC II): A Guide for Patient Management," *Phys. Ther.*, vol. 83, no. 5, pp. 455–470, May 2003, doi: 10.1093/ptj/83.5.455.
- [4] M. Kleynen, A. Moser, F. A. Haarsma, A. J. Beurskens, and S. M. Braun, "Physiotherapists use a great variety of motor learning options in neurological rehabilitation, from which they choose through an iterative process: a retrospective think-aloud study," *Disabil. Rehabil.*, vol. 39, no. 17, pp. 1729–1737, Aug. 2017, doi: 10.1080/09638288.2016.1207111.
- [5] M. K. Farlie, E. Molloy, J. L. Keating, and T. P. Haines, "Clinical Markers of the Intensity of Balance Challenge: Observational Study of Older Adult Responses to Balance Tasks," *Phys. Ther.*, vol. 96, no. 3, pp. 313–323, Mar. 2016, doi: 10.2522/ptj.20140524.
- [6] K. Holmqvist, M. Nyström, R. Andersson, and R. Dewhurst, *Eye tracking: a comprehensive guide to methods and measures*. Oxford ; New York: Oxford University Press, 2011.
- [7] A. Davies et al., "Exploring the Relationship Between Eye Movements and Electrocardiogram Interpretation Accuracy," *Sci. Rep.*, vol. 6, Dec. 2016, doi: 10.1038/srep38227.
- [8] M. Sibbald, A. B. H. de Bruin, E. Yu, and J. J. G. van Merriënboer, "Why verifying diagnostic decisions with a checklist can help: insights from eye tracking," *Adv. Health Sci. Educ.*, vol. 20, no. 4, pp. 1053–1060, Oct. 2015, doi: 10.1007/s10459-015-9585-1.
- [9] T. Balslev et al., "Visual expertise in paediatric neurology," *Eur. J. Paediatr. Neurol.*, vol. 16, no. 2, pp. 161–166, Mar. 2012, doi: 10.1016/j.ejpn.2011.07.004.
- [10] D. J. Manning, S. C. Ethell, and T. Donovan, "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph," *Br. J. Radiol.*, vol. 77, pp. 231–235, 2004, doi: 10.1259/bjr/28883951.
- [11] A. Desvergez, A. Winer, J.-B. Gouyon, and M. Descouts, "An observational study using eye tracking to assess resident and senior anesthetists' situation awareness and visual perception in postpartum hemorrhage high fidelity simulation," *PLOS ONE*, vol. 14, no. 8, p. e0221515, Aug. 2019, doi: 10.1371/journal.pone.0221515.
- [12] Z. Liu et al., "Analysis of Eye Tracking of Physiotherapist during Walk Rehabilitation," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, Oct. 2019, pp. 6832–6837. doi: 10.1109/IECON.2019.8926863.
- [13] K. McDuff et al., "Analyzing the Eye Gaze Behaviour of Students and Experienced Physiotherapists during Observational Movement Analysis," *Physiother. Can.*, vol. 73, no. 2, pp. 129–135, Spring 2021, doi: 10.3138/ptc-2019-0047.
- [14] K. Hayashi, S. Aono, M. Fujiwara, Y. Shiro, and T. Ushida, "Difference in eye movements during gait analysis between professionals and trainees," *PLOS ONE*, vol. 15, no. 4, p. e0232246, Apr. 2020, doi: 10.1371/journal.pone.0232246.
- [15] T. Bao, B. N. Klatt, S. L. Whitney, K. H. Sienko, and J. Wiens, "Automatically Evaluating Balance: A Machine Learning Approach," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 27, no. 2, pp. 179–186, Feb. 2019, doi: 10.1109/TNSRE.2019.2891000.
- [16] B. Klatt et al., "A Conceptual Framework for the Progression of Balance Exercises in Persons with Balance and Vestibular Disorders," *Phys. Med. Rehabil. Int.*, vol. 2, no. 4, 2015, Accessed: Nov. 20, 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4968039/>
- [17] J. Ferris, J. Zwier, W. J. Carender, and K. H. Sienko, "Differences between physical therapist ratings, self-ratings, and posturographic measures when assessing static balance exercise intensity," *Front. Rehabil. Sci.*, vol. 4, 2023, Accessed: Jan. 04, 2024. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fresc.2023.1096171>
- [18] A. Olsen, "The Tobii I-VT Fixation Filter: Algorithm description," Tobii Technology, 2012.
- [19] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates, 1988.
- [20] G. R. Loftus and N. H. Mackworth, "Cognitive determinants of fixation location during picture viewing," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 4, no. 4, pp. 565–572, Nov. 1978, doi: 10.1037/0096-1523.4.4.565.
- [21] B. E. Maki and W. E. McIlroy, "Control of rapid limb movements for balance recovery: age-related changes and implications for fall prevention," *Age Ageing*, vol. 35 Suppl 2, pp. ii12–ii18, Sep. 2006, doi: 10.1093/ageing/af1078.