33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Statistical Analysis of Gait Maturation in Children Based on Probability Density Functions

Yunfeng Wu*, *Member, IEEE*, Zhangting Zhong, *Student Member, IEEE*, Meng Lu, and Jia He

***Abstract*— Analysis of gait patterns in children is useful for the study of maturation of locomotor control. In this paper, we utilized the Parzen-window method to estimate the probability density functions (PDFs) of the stride interval for 50 children. With the estimated PDFs, the statistical measures, i.e., averaged stride interval (ASI), variation of stride interval (VSI), PDF skewness (SK), and PDF kurtosis (KU), were computed for the gait maturation in three age groups (aged 3–5 years, 6–8 years, and 10–14 years) of young children. The results indicated that the ASI and VSI values are significantly different between the three age groups. The VSI is decreased rapidly until 8 years of age, and then continues to be decreased at a slower rate. The SK values of the PDFs for all of the three age groups are positive, which shows a slight imbalance in the stride interval distribution within each age group. In addition, the decrease of the KU values of the PDFs is age-dependent, which suggests the effects of the musculo-skeletal growth on the gait maturation in young children.**

I. Introduction

The study of the early development of human mobility helps physiologists and neuroscientists understand the maturation of children's gait. According to Sutherland [1], an infant is able to sit upright at about 6 months after birth, begins to crawl after 9 months, and walks with immature control of posture at around 12 months. The gait in young children about 3 years old will become relatively mature, with a more stable walking pattern [2]. The findings of Beck *et al.* [3] suggested that the temporal and distance parameters in children were fixed by the age of 4 years. Norlin *et al.* [4], however, included a sample of 230 individuals from 3 to 16 years, and reported that the gait had not matured by the age of 8 years. Nonetheless, a key unanswered question is whether subtle changes in gait unsteadiness and stride-to-stride dynamics also occur in adolescents.

When young children first learn to walk, immature control of posture and gait results in large stride-to-stride fluctuations [5]. Some studies reported that walking variability decreases between childhood and adulthood [6], [7]. Hausdorff *et al.* [6] applied the fractal analysis method to study the stride-to-stride change in adolescents, and reported that the temporal structure of stride dynamics was associated with long-range, fractal organization. As Shumway-Cook and Woollacott suggested [8], analysis of the stride dynamics may provide a window into the development of neuromuscular control in children, but further quantitative studies still call for more computational tools to describe the mature of gait in children. The aim of the present study was to characterize the development of mature stride dynamics in young children, using statistical parameters and probability density functions.

Manuscript received April 16, 2011; accepted June 9, 2011. This work was supported in part by the Fundamental Research Funds for the Central Universities of China under grant No. 2010121061. *Asterisk indicates corresponding author.*

Y. F. Wu, Z. T. Zhong, M. Lu, and J. He are with Department of Communication Engineering, School of Information Science and Technology, Xiamen University, 422 Si Ming South Road, Xiamen, Fujian, 361005, China. Email: y.wu@ieee.org

II. Methods

A. Subjects

The gait database was obtained from the web page of PhysioNet [9]. There were 50 healthy children participants (aged 3–14 years, 25 boys and 25 girls) recruited from the local community in Boston, MA, USA [6]. The numbers of the subjects are 14, 21, and 15, for the 3- to 5-year-old, 6- to 8-year-old, and 10- to 14-year-old age groups, respectively. The children's parents were requested to provide informed written consent. None of these children had a history of neurological, cardiovascular, or musculoskeletal disorders [6].

B. Experiment Protocol

According to Hausdorff *et al.* [6], the children subjects were instructed to walk at their normal pace around a 400-m running track for 8 min. An investigator walked slightly behind each subject during the ambulation. Two ultrathin pressure-sensitive switches [10] were placed in each subject's right shoe, in order to record the force applied to the ground. The temporal signals were digitized by an on-board analog-to-digital converter with the sampling rate of 300 Hz and 12-bit resolution per sample, and then stored in a recorder (dimensions: $5.5 \times 2 \times 9$ cm; weight: 0.1 kg). The recorder was worn on the ankle cuff of each foot and held in place with a wallet on the ankle. The time series of stride interval were obtained with the pre-processing algorithm proposed by Hausdorff *et al.* [10].

We excluded the samples of the stride interval recorded in the first 1 min and the last 5 s, to minimize the start-up or ending effects of walking posture. A median filter [11] was applied to detect the outliers that were 3 standard deviations (SDs) in amplitude greater than the median value in the time series of stride interval. According to the well-known "three-sigma rule" [12], about 99.7% of the normally distributed probability values lies within 3-SD distance from the mean, which implies the outliers only occur with a small probability. The outliers detected, together with one stride before or after the outliers, were considered to be associated with the pauses during the gait monitoring. Therefore, we removed these outlier samples of stride interval before the gait analysis.

Figure 1 illustrates the outlier-free stride-to-stride interval time series of the children in three different age groups. Because we have eliminated the start-up effects, the first stride in each time series was made from the second minute in the 8-min monitoring.

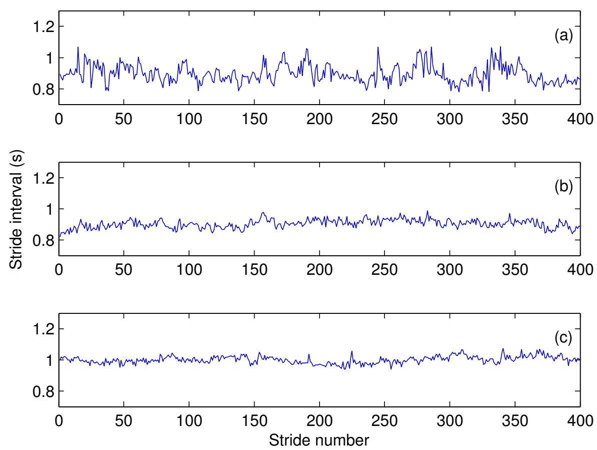


Fig. 1. Outlier-free time series of stride interval of the children (a) aged 47 months (in the group aged 3–5 years), (b) aged 82 months (in the group aged 6–8 years), and (c) aged 130 months (in the group aged 10–14 years), respectively. The first strides in (a)–(c) start after the start-up 1 min.

C. Gait analysis

1) Probability density function (PDF) estimation: In the present study, we first established the histogram as a PDF reference for each outlier-free stride interval time series. The histogram of stride interval was computed with a total of B bins, which could be used to calculate the probability of occurrence with B containers of equal length in the amplitude range. By following the choice of Scott's [13], the optimal number of bins that can minimize the mean squared error between the estimated histogram and the Gaussian density function was obtained as

$$B = \lceil \frac{(g_{max} - g_{min})}{3.49\, s\, n^{-1/3}} \rceil, \tag{1}$$

where s and n represent the SD and the number of samples in the stride interval time series, respectively; the highest and lowest values of stride interval g are denoted as g_{max} and g_{min}, respectively; and the operator $\lceil \cdot \rceil$ rounds the number of bins toward the nearest integer greater than or equal to it.

For each subject, we used the Parzen-window method [14] to estimate the PDF of stride interval from the outlier-free time series. Given a M-length stride interval time series, $\{g_m\}$, $m = 1, 2, \cdots, M$, the estimated PDF $\hat{p}(g)$ can be expressed as [15]

$$\hat{p}(g) = \frac{1}{M} \sum_{m=1}^{M} \omega\,(g - g_m), \tag{2}$$

where $\omega(\cdot)$ is a window function that integrates to unity. The Gaussian window function was used to estimate the Parzen-window PDF of stride interval, i.e.,

$$\omega(g - g_m) = \frac{1}{\sigma_{\mathrm{P}} \sqrt{2\pi}} \exp\left[\frac{-(g - g_m)^2}{2\sigma_{\mathrm{P}}^2} \right], \tag{3}$$

where σ_{P} represents the spread parameter that determines the width of a Gaussian window, the center of which is located at g_m [16], [17]. In order to determine the optimal spread parameter, the Parzen-window PDF was arranged with the same resolution as the histogram, i.e., the estimated probability density, $\hat{p}(g^b)$, $b = 1, 2, \cdots, B$, was also represented with B bins. Then the optimal spread parameter can be obtained by minimizing the mean-squared error (MSE) between the Parzen-window PDF, $\hat{p}(g^b)$, and the histogram, $\hat{h}(g^b)$, i.e., $\min \left\{ \frac{1}{B} \sum_{b=1}^{B} [\hat{p}(g^b) - \hat{h}(g^b)]^2 \right\}$. The optimal value of σ_{P} was set to be 0.01 in accordance with the minimization of MSE criterion.

2) Statistical parameters: There were four statistical parameters, in particular, averaged stride interval (ASI), variation of stride interval (VSI), skewness (SK), and kurtosis (KU), computed using the Parzen-window PDF estimated. The ASI and VSI are the mean and SD values of stride interval, i.e.,

$$\mu = \sum_{b=1}^{B} g^b \hat{p}(g^b), \tag{4}$$

and

$$\sigma = \sqrt{\sum_{b=1}^{B} (g^b - \mu)^2 \hat{p}(g^b)}. \tag{5}$$

The SK and KU are two parameters that usually measure the asymmetry and peakedness of the PDF [18]. The SK and KU can be computed from the moments of the PDF as

$$\mathrm{SK} = \frac{m_3}{(m_2)^{3/2}}, \tag{6}$$

and

$$\mathrm{KU} = \frac{m_4}{(m_2)^2}, \tag{7}$$

where m_j represents the jth central moment of the PDF, defined as [19]

$$m_j = \sum_{b=1}^{B} \left(g^b - \mu \right)^j \hat{p}(g^b). \tag{8}$$

III. Results

As shown in Figure 2, the degree of stride fluctuations is highest in the youngest child (aged 3–5 years), whereas the stride-to-stride variability of the other two children (aged 6–8 years and 10–14 years) becomes much smaller. It can be observed that the spread of the PDF of the child aged 47 months is much wider than those of the older children aged 82 and 130 months, respectively. The mean value of the stride interval was 0.897 s for the children aged 130 months, the amplitude of which was larger than those of the younger children (0.902 and 1.001 s for aged 47 and 82 months, respectively).

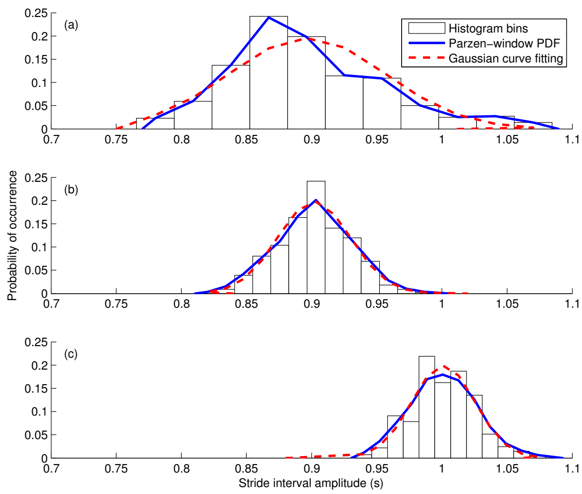


Fig. 2. Histograms and Parzen-window probability density functions (PDFs) estimated for the stride interval time series in Figure 1: of the children (a) aged 47 months (in the group aged 3–5 years), (b) aged 82 months (in the group aged 6–8 years), and (c) aged 130 months (in the group aged 10–14 years), respectively. Gaussian PDFs (*dashed curves*) are fit with the Gaussian distributions, the mean and variance parameters of which are equal to those of the corresponding histograms.

The values of the ASI and VSI computed from the Parzen-window PDFs for the children in different age groups are provided in Figure 3. We may observe that both the ASI and VSI values are age-dependent. The ASI value increases when the children grow up, whereas the VSI value decreases with age. The ASI value for the 6- to 8-year-old age group is about 0.056 s higher (p-value < 0.01) than that for the 3- to 5-year-old age group, and the ASI value is increased by 0.099 s (p-value < 0.0001) comparing the 10- to 14-year-old children with that for the 6- to 8-year-old children. In the meanwhile, the VSI is decreased by 0.023 s (p-value < 0.0001) in the children aged from 3 years to 8 years, and then continues to be decreased by 0.008 s (p-value < 0.001) until the children are 14 years old. Such results suggested that the children are more and more skilled to modulate large strides during the course of musculo-skeletal growth, and the ability to control stable strides is significantly improved in the children aged 3–8 years. It can therefore be inferred that the locomotor control system in children aged 3–8 years is still rapidly developing, and will reach maturity until they are 14 years old, when their gait patterns become very close to those of healthy adults [6].

The SK and KU results were shown in Figure 4. The SK values computed from both the histograms and Parzen-window PDFs were positive for the three age groups. The right-skewed PDFs (positive SK values) indicated that the mass of the distribution is located on the left side of the figure, which implies that more than a half of the stride interval samples are lower in amplitude than the mean of the PDF. It is also worth noting that the SK values first decreased rapidly from 0.31 (3–5 years old) to 0.14 (6–

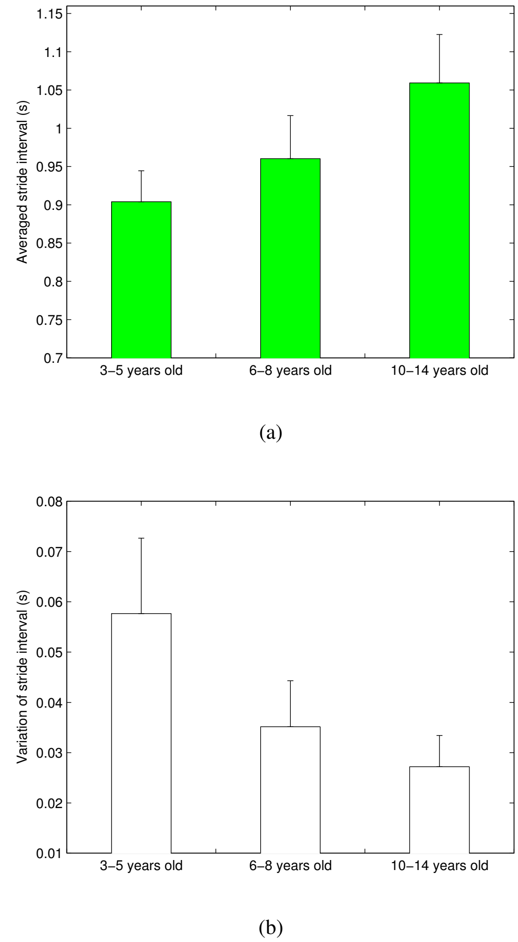
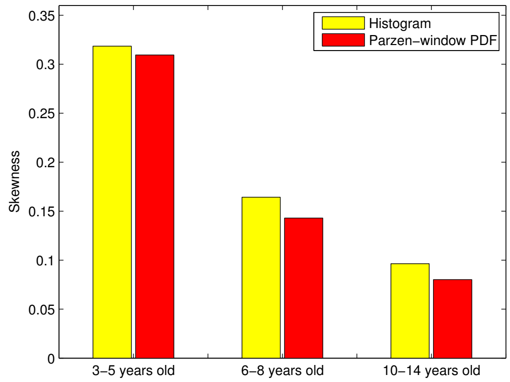


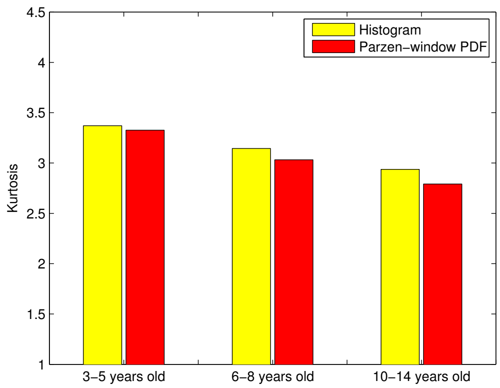
Fig. 3. Bar graphics of (a) averaged stride interval (ASI) and (b) variation stride interval (VSI) computed from the Parzen-window probability density functions (PDFs) of the children in the groups aged 3–5 years, 6–8 years, and 10–14 years, respectively. Vertical lines on the tops of the bars denote the standard deviation (SD) values. The values (mean $\pm$ SD) of the bars for the children of 3–5 years old (ASI: 0.904 ± 0.041 s, VSI: 0.058 ± 0.015 s), of 6–8 years old (ASI: 0.96 ± 0.056 s, VSI: 0.035 ± 0.009 s), and of 10–14 years old (ASI: 1.059 ± 0.063 s, VSI: 0.027 ± 0.006 s).

8 years old), and then continued to decrease at a slower rate until 14 years of age. It can be inferred that the stride interval distribution was imbalanced within an age group, but such an imbalance would be meliorated when the children grow up. The SK results demostrated our inference about the development process of the locomotor control system in young children described above.

The KU results are provided in Figure 4 (b). It can be observed that the KU values of the PDFs are higher than 3.0 (the typical value of a Gaussian PDF) when the children are 3–5 years old, are very close to 3.0 when the children are 6–8 years old, and are between 2.41 (the typical value of a raised cosine PDF) and 3.0 when the children are 10–14 years old. Such results indicated that the PDF curves become smoother and smoother during the course of musculo-skeletal growth in children, and the distributions of the stride-to-stride interval in children are not heavily concentrated on the means of the corresponding PDFs. It may be inferred that the musculo-skeletal growth enable the children to better control the strides at different speeds.



(a)



(b)

Fig. 4. Bar graphics of the mean values of (a) skewness (SK) and (b) kurtosis (KU) computed from the histograms and the Parzen-window probability density functions (PDFs) of the 3- to 5-year-old, 6- to 8-year-old, and 10- to 14-year-old age groups, respectively.

IV. CONCLUSION

According to the results obtained in the present study, we found that the gait continued to change up until at least the age of 14 years. The mean of stride interval was significantly increased with age, from 0.904 s in the youngest group to 1.059 s in the oldest group. The gait variability, in terms of VSI, is decreased rapidly until 8 years of age, and then continues to be decreased at a slower rate. The SK results indicated an imbalance in the stride interval distribution in young children, but could be meliorated later. Further results showed that the KU values will decrease with age, and such observations suggested that the musculo-skeletal growth enables the children to modulate a gait cadence with ease. The SK and KU results demonstrate that the Parzen-window PDF estimation method is also useful to quantitatively characterize the gait maturation in young children. The statistical parameters obtained in the present study could also be regarded as the dominant features for the pattern analysis of gait maturation. The future work will focus on the applications of computational intelligence techniques for the automatic analysis of gait patterns in young children [20].

REFERENCES

[1] D. Sutherland, "The development of mature gait," *Gait & Posture*, vol. 6, no. 2, pp. 163–170, 1997.

[2] D. Sutherland, R. A. Olshen, E. N. Biden, and M. P. Wyatt, *The Development of Mature Walking*. Oxford, UK: MacKeith, 1988.

[3] R. J. Beck, T. P. Andriacchi, K. N. Kuo, R. W. Fermier, and J. O. Galante, "Changes in the gait patterns of growing children," *Journal of Bone and Joint Surgery-American Volume*, vol. 63, no. 9, pp. 1452–1457, 1981.

[4] R. Norlin, P. Odenrich, and B. Sandlund, "Development of gait in the normal child," *Journal of Pediatric Orthopaedics*, vol. 1, no. 3, pp. 261–266, 1981.

[5] S. J. Hillmana, B. W. Stansfieldb, A. M. Richardsonc, and J. E. Robb, "Development of temporal and distance parameters of gait in normal children," *Gait & Posture*, vol. 29, no. 1, pp. 81–85, 2009.

[6] J. M. Hausdorff, L. Zemany, C. K. Peng, and A. L. Goldberger, "Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children," *Journal of Applied Physiology*, vol. 86, no. 3, pp. 1040–1047, 1999.

[7] K. G. Holt, E. Saltzman, C. L. Ho, and B. D. Ulrich, "Scaling of dynamics in the earliest stages of walking," *Physical Therapy*, vol. 87, no. 11, pp. 1458–1467, 2007.

[8] A. Shumway-Cook and M. H. Woollacott, *Motor Control: Theory and Practical Applications*, 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2000.

[9] G. B. Moody, R. G. Mark, and A. L. Goldberger, "PhysioNet: a web-based resource for the study of physiologic signals," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 70–75, 2001.

[10] J. M. Hausdorff, Z. Ladin, and J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *Journal of Biomechanics*, vol. 28, no. 3, pp. 347–351, 1995.

[11] Y. F. Wu and S. Krishnan, "Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis," *Medical & Biological Engineering & Computing*, vol. 47, no. 11, pp. 1165–1171, 2009.

[12] G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering*. Hoboken, NJ: Wiley, 1994.

[13] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.

[14] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[15] R. M. Rangayyan and Y. F. Wu, "Screening of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 53–58, 2010.

[16] Y. F. Wu and S. Krishnan, "Statistical analysis of gait rhythm in patients with Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 150–158, 2010.

[17] Y. F. Wu and L. Shi, "Analysis of altered gait rhythm in amyotrophic lateral sclerosis based on nonparametric probability density function estimation," *Medical Engineering & Physics*, vol. 33, no. 3, pp. 347–355, 2011.

[18] R. M. Rangayyan and Y. F. Wu, "Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions," *Medical & Biological Engineering & Computing*, vol. 46, no. 3, pp. 223–232, 2008.

[19] J. P. Marques de Sa, *Applied Statistics using SPSS, STATISTICA, and MATLAB*. Berlin, Germany: Springer-Verlag, 2003.

[20] Y. F. Wu and S. Krishnan, "Combining least-squares support vector machines for classification of biomedical signals: a case study with knee-joint vibroarthrographic signals," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 23, no. 1, pp. 63–77, 2011.