Quantifying the performance benefits of middle distance running super spikes

Montgomery Bertschy[1*], Victor Rodrigo-Carranza[1,2*], Ethan W.C. Wilkie[3*], Laura A. Healey[4], Jeremy Noble[3], Wayne J. Albert[3], Wouter Hoogkamer[1]

[1] Department of Kinesiology, University of Massachusetts, Amherst, MA, USA

[2] Sports Performance Research Group (GIRD) University of Castilla-La Mancha, Toledo, Spain

[3] University of New Brunswick, Fredericton, NB, Canada

[4] PUMA SE, Innovation, Somerville, MA, USA

* These authors contributed equally to this work

Corresponding Author:

Wouter Hoogkamer

Integrative Locomotion Laboratory

Department of Kinesiology

University of Massachusetts, Amherst

Amherst, MA, 01003-9258 USA

Email: whoogkamer@umass.edu

ORCID:

Wouter Hoogkamer 0000-0003-0524-0196

Victor Rodrigo-Carranza 0000-0003-1637-7550

Laura Healey 0000-0003-0613-1629

Abstract

Background Quantifying the potential benefits of advanced footwear technology (AFT) track spikes in middle-distance events is challenging, because repeated maximal effort trials (as in sprinting) or aerobic running economy trials (as in long-distance running) are not feasible.

Objective We propose and evaluate a novel approach to assess the benefits of AFT spikes that consists of a series of 200m intervals at self-perceived middle-distance race pace.

Methods We conducted four experiments to evaluate the validity, sensitivity, reproducibility, and utility of this novel method.

Results (1) Participants ran 1.26% slower (p=0.0188) in spikes with 200g added mass than in control spikes, exactly equal to the known effects of shoe mass on running performance. (2) Participants ran significantly faster in AFT prototype spikes than in traditional spikes (p=0.0296). (3) Participants performed our interval-based method to compare two other AFT prototype spikes against traditional spikes, on three separate days. Group-level results were consistent across days, but to evaluate individual responses at least two separate sessions are needed. (4) Participants ran significantly faster in two AFT spike models than in traditional spikes (2.22% and 1.74%). Speed was similar between a third AFT spike model and the traditional spikes. These speed results were mirrored by changes in step length, as participants took significantly longer steps in the two faster AFT spike models (2.31% and 1.92%), while step length was similar between the other spikes.

Conclusion Our novel interval-based approach is a valid and reliable method to quantify differences between spikes at middle-distance running intensity.

Key points

- A novel, effort-based methodology allows for valid and reliable comparison of spikes for middle-distance running events.
- AFT track spikes are found to be significantly faster than traditional track spikes, primarily due to increases in step length.
- Group-level results were consistent across days, but at least two separate sessions are needed to evaluate individual responses.

Declarations

Compliance with Ethical Standards

Ethical approval The study was performed in accordance with the ethical standards of the Declaration of Helsinki. Ethics approval was obtained from the University of Massachusetts Institutional Review Board (Protocol# 2647), from the University of New Brunswick Research Ethics Board (REB #2021-079) and the University of Castilla-La Mancha University Ethics Committee (Protocol CEIC924)

Informed consent Written informed consent was obtained from all individual participants included in the study.

Funding

This research was partly supported by a research contract from PUMA SE with the University of Massachusetts, Amherst. All spikes were provided by PUMA SE.

Conflicts of interest / Competing interests

Montgomery Bertschy, Victor Rodrigo-Carranza, Ethan W.C. Wilkie, Jeremy Noble and Wayne J. Albert have no conflicts of interest relevant to the content of this article. Wouter Hoogkamer is a paid consultant to PUMA and has received research grants from PUMA and Saucony. Laura A. Healey is an employee of PUMA. No footwear company had any influence on the conceptualization of this study or results presented in this publication.

Data availability

All data discussed in this manuscript are provided in the tables.

Authors' contributions

Conceptualization: Wouter Hoogkamer; Experimental design: Montgomery Bertschy, Laura A. Healey, Wouter Hoogkamer; Data collection: Montgomery Bertschy, Victor Rodrigo-Carranza, Ethan W.C. Wilkie, Jeremy Noble; Data analysis: Montgomery Bertschy, Victor Rodrigo-Carranza, Ethan W.C. Wilkie, Jeremy Noble, Wayne J. Albert; Writing - original draft preparation: Montgomery Bertschy, Victor Rodrigo-Carranza, Ethan W.C. Wilkie, Laura A. Healey, Wouter Hoogkamer; Writing - review and editing: Montgomery Bertschy, Victor Rodrigo-Carranza, Ethan W.C. Wilkie, Laura A. Healey, Jeremy Noble, Wayne J. Albert, Wouter Hoogkamer

1 Introduction

Once thought to be the last unbreakable barrier on the track [1], the magical sub-4-minute mile mark stood strong for decades, until Sir Roger Bannister in 1954 completed the unthinkable and ran the 1609 meters in 3:59.4 seconds. In February 2023, 52 male professional, club, and collegiate level athletes broke the 4-minute mile barrier all within the span of an hour at Boston University's indoor track facility. A similar phenomenon was observed in the women's section, as three athletes broke the 4:30 mile barrier, and 17 ran sub-4:36, a mark more equivalent to the men's barrier.[2] Over the past decades, the number of athletes breaking these barriers had been increasing gradually but there has been a substantial increase since 2021.While the Boston University track has been known to be fast, another important common feature among these performances was the use of advanced footwear technology (AFT) track spikes, colloquially known as "super spikes". Similar to the now common AFT marathon racing shoes (e.g., Nike Vaporfly), AFT track spikes feature a lightweight, compliant, and resilient midsole foam, often combined with a stiff midsole plate. While we have argued before that anecdotal performances should not be used as evidence on the efficacy of specific shoes [3], with the widespread improvement across middle distance running performances, this warrants a closer look.

The recent improvements on the track are similar to improvements in road running performances since 2017 which have been attributed to advances in road running shoes.[4–6] While the adoption of AFT road shoes happened in parallel with the publication of a randomized crossover study showing 4% improvements in running economy, scientific evidence on potential benefits of AFT track spikes is currently limited to race results analyses (preprint: [7]) and preliminary laboratory data.[8] Willwacher et al., [7] found

in an observational study that AFT track spikes had a greater impact on performance in middle-distance events (0.31 - 1.33%) than in sprint events (0.22 - 0.66%) for both men and women.

Traditionally, the effects of footwear on middle distance running performance have been studied in the lab by measuring running economy, or on the track with time trials. Barnes and Kilding [9] observed that running economy in traditional lightweight track spikes was ~3% worse than in AFT road shoes during treadmill running at speeds between 3.9 and 5.0 m/s. Oehlert et al.,[8] showed that different AFT track spikes improved running economy by ~2% compared to a traditional spike model during treadmill running at 14 km/h in women and at 16 km/h in men. Lab-measured changes in running economy have been shown to translate to changes in running performance [10], but for speeds faster than 3 m/s, performance improvements can be expected to be smaller than observed improvements in running economy.[11] Furthermore, running economy measures are only valid when quantified during running at a sustainable, aerobic, steady-state intensity. For very fit, elite participants, this could involve running as fast as 5.8 m/s.[12,13] However, world record pace for middle distance events ranges from 5.9 m/s (women's 5000m) to 7.9 m/s (men's 800m). How footwear affects running economy in high-caliber runners at 5.0 m/s [14] might not be indicative of how this footwear affects running performance at a speed of 7.9 m/s, almost 60% faster. At these high middle-distance running speeds, at intensities above aerobic steady-state capacity, quantifying physiological responses to different spikes is not feasible or valid (see [3]).

An alternative approach to quantify the benefits of spikes could be to have athletes perform a series of time trials in different spikes. Indeed, we showed earlier that

3,000m time trial performances are slower in shoes with added mass.[10] Participants ran three time trials on an indoor track, each at the same time of day at the same day of the week, a week apart. Performing race-distance time trials at maximal effort across multiple days can seriously interfere with training programming, while doing multiple high effort time trials on a single day, on the other hand, will affect the validity of the results through fatigue building up.

In this paper we propose, evaluate, and use a novel approach to quantify potential benefits of AFT middle-distance spikes. Rather than comparing time trial performance using race distance time trials at maximal race effort over multiple days, we asked athletes to run a series of 200m intervals at self-perceived race pace (800m or 1500m, based on their specialty), within a single session. Series of 200m intervals at race pace are a common component of middle-distance training, [15,16] and all participants were familiar with this type of workout. This is important as the premise of the proposed protocol is to measure differences in performance, while controlling effort. In a series of experiments, we compared up to four different spike models, while participants ran at least two intervals in each model. First, we evaluated the validity of our novel approach by comparing traditional spike models that were nearly identical and only differed in mass (control vs. control + 200g) by verifying the interval time differences against the theoretically expected differences. Second, we assessed the sensitivity of this novel protocol in comparing new prototype AFT spikes against traditional spikes. Third, we evaluated the reliability between three different days for two AFT spikes against traditional spikes, and finally we used this protocol to compare several commercially available AFT spikes against traditional spikes. We used an inertial measurement unit

(IMU) placed at the pelvis to assess if any differences in running speed were predominantly driven by changes in step frequency or by changes in step length.

2. Materials and Methods

2.1 A Series of Experiments

This study involved four separate experiments. The inclusion criteria were generally similar for all four experiments: fitting US women's size 7, 8.5 or 9, or US men's size 9 or 11 track spikes (Experiment 1) or fitting US men's size 9 track spikes (Experiments 2, 3 and 4); being 16 years of age or older (Experiment 1 and 2) or 19 years of age or older (Experiment 3 and 4); currently participating in track and field training; having at minimum two years of experience training in track spikes; being free of any orthopedic, cardiovascular, or neuromuscular conditions; and having not sustained injury nor undergone any surgery within the 3 months prior to participation. Recruitment of participants was done through means of personal contacts, local track clubs, and local university track teams. Participants were Tier 3: highly trained, national level, but not elite,[17] with average personal bests of 2:02min in 800 to 8:55min in 3000m across the different experiments (average IAAF scores for each experiment: 815, 762, 817, 761, respectively).[18] Because of the different timelines of the experiments and the increasing availability of AFT spikes with time, participants experience running with AFT varied between experiments: all participant had experience running in spikes, for Experiment 2 only two participants had run in AFT road shoes, none in AFT spikes, for Experiments 3 and 4 all participants had experience running with AFT shoes and/or spikes (Experiment 1 only used traditional spikes). The characteristics of the participants in each experiment

are shown in Table 1. The experiments were performed in accordance with the ethical standards of the Declaration of Helsinki. Ethics approval was obtained from the University of Massachusetts, Amherst Institutional Review Board (Protocol # 2647), the University of New Brunswick, Fredericton Research Ethics Board (REB # 2021-079), and the University of Castilla-La Mancha University Ethics Committee (Protocol CEIC924). Before taking part in the study, participants provided informed written consent.

2.1.1 Spike Conditions

For all experiments we used the PUMA EvoSpeed Distance 9 (TRAD) as the control condition; a typical traditional track spike, lightweight in construction, while providing only minimal cushioning and no added bending stiffness elements. It has a low-profile EVA based foam with a flat geometry and a stack height of 17mm (Fig 1). In Experiment 1, we compared the standard version of this spike (TRAD, 137-169g) against the same spikes with 200g of added mass (TRAD+, 331-368g) (Table 2).

In Experiment 2 we compared two track spike conditions: TRAD and a PUMA AFT prototype middle-distance spike (AFTP1, 179g). The AFTP1 spike combines a compliant, highly resilient midsole foam with an embedded carbon fiber plate. It features PUMA's Nitro foam, a lightweight PEBA-based foam compound, forming a thicker, rockered midsole with midsole thickness of 19mm (Fig 1; Table 2).

In Experiment 3 we compared three track spike conditions: TRAD, a PUMA AFT prototype long-distance spike (AFTP2, 151g) and a PUMA AFT prototype middle-distance spike (AFTP3, 149g, different from AFTP1). The AFTP2 spike features

PUMA's Nitro foam, a lightweight PEBA-based foam compound, forming a rockered midsole with midsole thickness of 18mm, but no carbon fiber plate. The AFTP3 spike combines a compliant, highly resilient midsole foam with an embedded carbon fiber plate. It features PUMA's Nitro foam, a lightweight PEBA-based foam compound, forming a thicker, rockered midsole with midsole thickness of 19mm (Fig 1; Table 2).

In Experiment 4 we compared four commercially available track spike conditions: TRAD, a commercially available PUMA AFT spike (AFTC1) and two competitor brand AFT spikes (AFTC2 and AFTC3). The AFTC1 was the PUMA evoSPEED Long Distance Nitro Elite+ (154g). It features PUMA's Nitro foam, a lightweight PEBA-based foam compound, forming a rockered midsole with midsole thickness of 18mm, but no carbon fiber plate. The AFTC2 (140g) and AFTC3 (163g) were commercially available competitor brand AFT spike models, featuring similar, but different, AFT properties such as increased compliance, resilience, bending stiffness, stack height and rocker geometry (Fig 1; Table 2).

2.1.3 Experiment 1 - Validation

To evaluate the external validity of our novel approach, we compared 200m interval time in control spikes versus spikes that were nearly identical and only differed in mass (control vs. control + 200g). Similar to Hoogkamer et al., [10], we used a mass intervention to predictably induce a change in running energetics.[19,20] Frederick et al., [19] showed that, across a range of speed, adding 100g mass to each shoe increases the energy cost of running by 1%. By how much a 1% increase in energy cost will slow down a runner depends on how fast they are running. Kipp et al., [11] provided a theoretical

framework supported by observations in Hoogkamer et al., [10] that predicts that when running at 6.67m/s (200m in 30s), a 1% increase in energy cost will lead to a 0.6% slower interval. Therefore, with 200g added to each spike, we hypothesized that participants would run 1.2% slower.

2.1.4 Experiment 2 - Sensitivity

Next, we evaluated if this approach had the sensitivity to find differences in what may be seen as "anecdotal" benefits of AFT spikes over traditional spikes. For this newly developed prototype middle distance spike, we anticipated that the metabolic savings from the modern midsole foam and the embedded plate would add up to about 2%, based on preliminary running economy results from Oehlert et al., [8]. Therefore, we hypothesized that participants would run 1.2% faster in the prototype spikes.

2.1.5 Experiment 3 – Reliability

We then evaluated the reliability of our novel approach to quantify the potential benefits of different AFT spikes technologies. The objective was to assess the consistency of the performance benefits from AFT spikes across different days. We hypothesized that different AFT spikes would result in faster times than TRAD, similarly and consistently across the 3 measurement days.

2.1.6 Experiment 4 - Commercially Available Spikes

Finally, we used our novel approach to quantify potential benefits of a set of different commercially available AFT spikes over traditional spikes. The aim was to

evaluate if potential performance gains from AFT spikes differed between models. We hypothesized that the different AFT spikes all would result in 1.2% faster times than TRAD, with no detectable differences between AFT models. Further, we hypothesized that step frequency would be similar between conditions, while step length would be about 1% longer in the AFT spikes [14,21].

2.2 Generic Experimental Protocol

2.2.1 General protocol

A single-blinded crossover methodology was implemented as part of the protocol for Experiments 1, 2 and 4. During Experiment 3 the participants were able to see the spikes but were not aware of the spike properties. All experiments took place at an outdoor 400-meter track facility on days with little to no wind. The middle-distance track runners enrolled in the study performed a series of 200-meter intervals at self-perceived mile or 800m race pace. For middle-distance running we could not use maximal effort trials and the effectiveness of the protocol depended on the participants' ability to run every trial at a similar sub-maximal effort. To minimize bias, it is important that the participants are unaware of the spike intervention and that they are able to control their effort at a consistent submaximal level. Aiming to achieve the latter, we opted for 200m intervals at self-perceived race pace, since this is a common component of middle-distance training,[15,16] and all participants, like most middle-distance runners, were familiar with this type of workout. To minimize any confounding effects of participants running the first (excitement) or last ("emptying the tank") intervals at a higher effort, the number of intervals that the researchers told the participants to run was higher than the

number of intervals they would actually run. Their first interval was considered a habituation trial and not included in the evaluation, and then after running all the needed experimental intervals, the researchers informed the participants that the data collection was complete. Experiment 1 consisted of 6 experimental trials (3 intervals in both conditions), and we asked participants to run 8 intervals, but then informed them that the experiment was complete after 7 intervals. Similarly, for Experiments 2 and 4 we asked participants to run 10 intervals, but we only included 4 and 8 experimental trials, respectively (two intervals in each spike condition). For Experiment 3 the participants performed the same protocol on three separate visits, so they ran all 8 intervals, but we only analyzed the 6 experimental intervals.

The order of spike conditions was randomized for each participant prior to each session. Participants performed each 200m interval from a standing start (similar to a middle-distance race start). Participants were unaware of the times they were running for each interval as they were not using a watch and researchers did not disclose the times they recorded until after the data collection was complete. Rest between intervals was ten minutes, to allow for changing spikes, filling out questionnaires (Experiments 2 and 4), and walking to the start of the next interval.

2.2.2 Experimental Setup

When the participant arrived at the track facility, we explained the study protocol to them, and they signed all the necessary forms [i.e., the Get Active Questionnaire (CSEP, 2017) or the Physical Activity Readiness Questionnaire (ePARmed-X+) and informed consent document]. We asked participants to run a warmup similar to as they

would for a 200m interval session, followed by drills and stretches as they saw fit. For Experiment 4, after their warmup, we placed an inertial measurement unit (IMU; Vicon Blue Trident sensor, Vicon, Oxford, UK) at the participant's sacrum with a clip attached to the back of the waistband on their shorts.[22] Next, the participant removed their shoes and researchers put fabric spike blinders on the participant's legs (Experiments 1, 2 and 4). Then the researchers put the spikes on the participant's feet and tightened the laces. Following this, spike blinders were pulled down over the spikes. Spike blinders fit around the participants spikes and lower legs in a non-restrictive fashion, providing a negligible addition to spike mass.

At the 200m line the researcher started the IMU recording (Experiment 4) and then moved towards the finish line. For the timing of each interval, researchers started their timers when the participant started running, indicated by them quickly lowering one hand from a raised position, and stopped their stopwatches when the participant crossed the finish line. Once a trial was complete, the 200m time was recorded and the IMU capture was stopped (Experiment 4). This was repeated for all intervals.

2.3 Data Analyses and Statistics

We calculated running speed in m/s by dividing the 200m distance by the interval time. To calculate step frequency and step length in Experiment 4, we exported the raw IMU data from the Capture.U Desktop. Orientation of the IMU was verified from acceleration signal by finding the gravitational component (-9.81m/s^2) during quiet standing before the running trial began. The vertical component was filtered using a 5Hz cut-off frequency and multiplied by subject mass to create a vertical ground reaction

force estimation, as demonstrated by Day et al., [22]. Peaks in the vertical ground reaction force estimate were used to calculate step time and its inverse, step frequency. Average speed for each interval was calculated as interval distance (i.e., 200m) divided by interval time. Average step length was calculated as the product of average speed and average step time. Next, we averaged the values across the two intervals that were performed in the same spike condition.

We used linear-mixed effect models for each experiment [23]. For Experiments 1 and 2, we created models with spike as a fixed effect, subject as a random intercept, trial as a random effect, and speed as the output variable. For Experiment 3, the model had spike and day as fixed effects, subject as a random intercept, trial as a random effect, and speed as the output variable. Additionally, for each day of this experiment, we created a model with spike as a fixed effect, subject as a random intercept, trial as a random effect, and speed was the output variable to verify that the results did not change per day. For Experiment 4, we created three models with spike as a fixed effect, subject as a random intercept, and speed, step length, and step frequency as the output variables, respectively. A traditional level of significance ($\alpha = 0.05$) was employed. Tukey HSD tests were conducted for each pairwise comparison of spike conditions. Additionally, we calculated effect size (ES) as Cohen's d and reported percent differences for each measure with the intention of providing relevant context for our determined p values. ES with values of 0.2, 0.5, and above 0.8 were considered as small, medium, and large, respectively [24]. To evaluate the reliability, in Experiment 3 we calculated the coefficient of variation (CV = SD / mean × 100%) within each day and for each spike, across days. The ranges of values considered for categorizing CV% were as follows: CV < 10 as very good, 10–20 as good,

20–30 as acceptable, and CV > 30 as unacceptable [24]. Further, for each participant we calculated the average speeds in each spike across days 1 and 2, and across days 1, 2 and 3. Then, we calculated the speed difference between the TRAD and the AFTP3 spikes. Next, we calculated Pearson's correlation coefficients, for these differences, between each individual day and the 3-day average, and between the 2-day average and the 3-day average. Correlation coefficients with values of 0.1, 0.3, and 0.5 were considered as small, medium, and large, respectively, with 1.0 being a perfect correlation [24]. Statistical analyses were performed in R using lme4 package[25] for the linear mixed effect models and emmeans package[26] for post hoc Tukey HSD tests. Descriptive statistics are reported as mean ± standard deviation (SD). For visual display in the figures, we averaged the 200m interval times that were performed in the same spike condition for each experimental session (three trials for Experiment 1; two trials for Experiments 2, 3 and 4).

3 Results

In Experiment 1, participants ran significantly slower in the TRAD+ spikes (5.96 ± 0.53 m/s) than in the TRAD spikes (6.04 ± 0.50 m/s) (1.26%, ES = 0.16, p = 0.019; Fig. 2). In Experiment 2, participants ran significantly faster in the AFTP1 spikes (6.42 ± 0.41 m/s) than in the TRAD spikes (6.28 ± 0.49 m/s) (2.45%, ES = 0.31, p = 0.030; Fig. 2).

In Experiment 3 participants ran significantly faster in the AFTP2 spikes (6.52 ± 0.36 m/s) and the AFTP3 spikes (6.60 ± 0.36 m/s) than in the TRAD spikes (6.40 ± 0.34 m/s) (1.87%, ES=0.34, p < 0.0001 and 3.12%, ES= 0.57, p < 0.0001, respectively).

Running speeds in AFTP3 spikes were significantly faster than in AFTP2 spikes (1.22%, ES=0.19, p = 0.0004). The between-spike results were consistent between days. The CV% values were classified as very good for inter-session (5.53%, 5.44% and 5.69% for day 1, 2 and 3 respectively) and intra-session (5.29%, 5.53% and 5.46% for TRAD, AFTP2 and AFTP3 respectively). The linear mixed effects model analyses for each day separately all showed similar differences between spikes, both in magnitude (% change and effect size) and in significance (Table 3).

The speed difference between TRAD and AFTP3 for each individual participant varied across days. Pearson's correlation coefficient between day 1, day 2 and day 3, and the 3-day average was large: 0.85, 0.77, and 0.88, respectively. However, the Pearson's correlation coefficient between the 2-day average and the 3-day average was approaching a perfect correlation: 0.97.

In Experiment 4 participants ran significantly faster in the AFTC1 spikes (6.44 ± 0.46 m/s) and the AFTC2 spikes (6.41 ± 0.51 m/s) than in the TRAD spikes (6.30 ± 0.47 m/s) (2.22%, ES= 0.31, p = 0.002 and 1.74%, ES= 0.22, p = 0.035, respectively; Fig. 2). Running speeds in the AFTC3 spikes (6.36 ± 0.48 m/s) were not significantly different from the TRAD spikes (0.95%, ES=0.13, p = 0.493). Running speeds were not significantly different between the AFTC spikes (AFTC1 vs. AFTC2 p = 0.800; AFTC1 vs. AFTC3 p = 0.116; AFTC2 vs. AFTC3 p = 0.534), but the numerical difference between AFTC1 and AFTC3 was more than 1% (1.24%, ES = 0.17).

Step frequency was not different between spikes (all p > 0.52; Fig. 2), but participants ran with significantly longer steps in AFTC1 spikes (1.93 ± 0.15 m) and AFTC2 spikes (1.92 ± 0.16 m) than in the TRAD spikes (1.88 ± 0.15 m) (2.31%, ES =

0.33, p = 0.001; 1.92%, ES = 0.26, p = 0.008; Fig. 2). Step lengths in the AFTC3 spikes (1.91 ± 0.15 m) were not significantly different from the TRAD spikes (1.50%, ES = 0.20, p = 0.0512). Step lengths were not significantly different between the AFTC spikes (AFTC1 vs. AFTC2 p = 0.893; AFTC1 vs. AFTC3 p = 0.472; AFTC2 vs. AFTC3 p = 0.875).

4 Discussion

In this paper we proposed, evaluated and used a novel approach to quantify potential benefits of middle-distance AFT spikes. The combined results indicate that our novel interval-based approach has the validity, sensitivity and reliability to quantify the benefits of middle-distance AFT spikes. The results of Experiment 1 provide insights into the validity of our interval-based approach. On average, participants ran 1.26% slower in the spikes with 200g added mass, similar to the 1.2% difference we hypothesized. The results of Experiment 2 indicate that our approach has the sensitivity to confirm what so far has been considered "anecdotal" benefits of AFT spikes over traditional spikes. Participants ran significantly (2.45%) faster in AFT prototype spikes than in traditional spikes. The results of Experiment 3 indicate that our approach is reliable. On three separate days between spike differences in running speed were similar, both in magnitude (% change and effect size) and in significance. However, at the level of individual runners, the responses to AFT spikes varied between days. Our data suggests that two separate sessions are needed to obtain reliable individual data. Finally, we used our interval-based approach to compare several commercially available AFT spikes against traditional spikes. In two AFT spike models participants ran significantly (2.22% and

1.74%) faster than in the traditional spikes. Speed was similar between a third AFT spike model and the traditional spikes. These speed results were mirrored by changes in step length, as participants took significantly (2.31% and 1.92%) longer steps in the two faster AFT spike models, while step length was similar between the third AFT spike model and the traditional spikes.

4.1 Construct Validity

Shoe mass is one of the most studied features of running shoes [10,19,20,27]. There is currently a strong consensus that there is a linear relationship between the increase in shoe mass and the increase in running energy cost, with an increment in the oxygen/energy cost of 1% per added 100g of shoe mass. Kipp et al., [11] provided a theoretical framework to translate changes in energy cost into changes in running performance, based the known curvilinear relationship between running speed and energy cost and the cost of overcoming air resistance, and supported by shoe-mass time trial results from Hoogkamer et al., [10]. Using this framework, we anticipated that at 6.67m/s participants would run 1.2% slower in the TRAD+ spikes with 200g added mass. Our participants ran their intervals at 6.0m/s; at this speed the framework predicts a 1.26% slower speed in the TRAD+ spikes. We found that on average our participants ran 1.26% slower, confirming the construct validity of our interval-based approach.

4.2 Sensitivity

Next, we evaluated if our interval-based approach is sensitive enough to find differences between AFT spikes and traditional spikes. We anticipated that speed

differences between AFT spikes and traditional spikes would be similar to the effect of 200g added mass and we hypothesized that participants would run 1.2% faster in the AFT prototype spikes. Intriguingly, participants ran 2.45% faster in the AFT prototype spikes than in the traditional spikes, which is equivalent to the effect of 400g added mass.

4.3 Reliability

We evaluated the reliability of this protocol by asking 12 participants to run in the same 3 spike conditions with the same protocol on 3 separate days. We found that, at the group level, results were similar between all three days (Table 2) and similar to the 3-day average values (significant speed improvements of 1.87 and 3.12% of the AFTP2 and AFTP3, compared TRAD, respectively). The Pearson correlation coefficient for Day 1, 2, and 3 vs. the 3-day average was large: 0.85, 0.77, and 0.88, respectively. However, the correlation coefficient for 2-day vs. 3-day average results was substantially higher (0.97), suggesting that a minimum of 2 days is needed to quantify dependable individual responses to AFT spikes. Quantifying individual responses to AFT shoes is common when aiming to identify biomechanical mechanisms underlying improvements from AFT [28], or characteristics of runners who experience more or less benefits[29–31]. However, as discussed by Barrons et al., [32], such analyses critically depend on the accuracy of the response measurement. Note that both for running economy measures [32] and for speed outcomes from our interval-based approach, multiple sessions with repeated conditions are recommended, when attempting to quantify individual responses to AFT. Interestingly, even though the differences between spikes varied between days for some participants, the participants who ran faster in the AFTP3 spikes (with a curved carbon

fiber plate) than the AFTP2 spikes (no plate) did consistently (on each day) run faster in the AFTP3 spikes than in the TRAD spikes, while the participants who ran faster in the AFTP2 spikes than the AFTP3 spikes did consistently (on each day) run faster in the AFTP2 spikes than in the TRAD spikes (Fig 3). This suggests that when only interested in determining which spike is best suited for an individual, a single session in one day (with repeated trials) may be sufficient.

4.4 Quantifying the Benefits of Super Spikes

Finally, we used our novel approach to quantify potential benefits of a set of different commercially available AFT spikes over traditional spikes. Participants ran significantly faster in the AFTC1 (2.22%) and AFTC2 (1.74%) than in the TRAD spikes. However, running speeds in the AFTC3 spikes were not significantly different from the TRAD spikes (0.95%), nor between the AFTC spikes. The improvements from the commercially available AFT spikes compared to the traditional spikes were similar, but slightly less than those of the prototype AFT spikes evaluated in Experiments 2 and 3 (2.22 - 3.12%) of this study. Over 800 or 1500m at these speeds athletes would run 2.5 or 5 seconds faster, respectively, in these AFT spikes than in traditional spikes.

While the effects of spikes on sprinting performance have been studied for at least 50 years, with an inquiry into the potential benefits of the banned Puma brush spikes in the early seventies,[33] here we specifically focused on middle-distance running spikes. Given the novelty of AFT spikes, and the challenges of quantifying their potential benefits on middle-distance running performance,[3] we are not able to directly compare our findings to the literature.

Willwacher et al., [7] performed in a longitudinal, observational and retrospective study design evaluating the 100 best performances per year for men and women in outdoor track events from 2010 to 2022. They observed more pronounced improvements for the 1500m and longer distances longer (0.66 - 1.84%) than shorter distances (100m to 800m: 0.31-0.63%) for women, and less of an effect of AFT spikes in men. When comparing running economy results for AFT spikes at 4.4 and 5.0 m/s, Oehlert et al., [8] found 1.8 and 2.1% improvements for two AFT spike models as compared to traditional spikes. Such running economy improvements can be expected to improve middle-distance running performance by ~1.2% [11], which is less than the interval speed improvements we observed in the current study at 6.4 m/s. Although, not specific to AFT spikes, some earlier studies have evaluated 3,000m time trial performance on the track in AFT road shoes and reported improvements ranging from 1 to 5%, similar to our spike results [30,34].

Across Experiment 2, 3 and 4, participants consistently ran faster in the AFT spikes than in the traditional spikes (with the exception of AFTC3). The three experiments each used slightly different models of AFT spikes and the group-level benefits varied from 2.45% in Experiment 2, 1.87 and 3.12% in Experiment 3 to 2.22 and 1.74% in Experiment 4. The consistency of these findings, while these experiments were performed by different research teams in three different countries (US, Canada and Spain) is a testament of the robustness of our interval-based approach and further supports the validity of the observed benefits of AFT spikes.

Intriguingly, the faster speeds in the AFTC1 (2.22%) and AFTC2 (1.74%) spikes as compared to the TRAD spikes in Experiment 4, were accompanied by 2.31% (i.e.,

4.3cm) and 1.92% (i.e., 3.6cm) longer steps in the two faster AFTC1 and AFTC2 spikes, respectively. Step lengths were similar between the other spike models. Furthermore, step frequency was similar across all spikes (~3.35 Hz, or 201 steps per minute). Combined, this suggests that the performance benefits from AFT spikes in middle-distance running originate from the spikes allowing runners to take longer step lengths without requiring any compensation in step frequency. With average step lengths of 1.93m and 1.92m in the AFTC1 and AFTC2 spikes, the participants took only 103.6 steps and 104.1 steps during their 200m, respectively, about 2.5 steps less than in the traditional spikes. Over a 1500m race this adds up to 21 or 17 less steps in the AFTC1 and AFTC2 spikes than in traditional spikes, respectively.

These biomechanical changes are similar to those reported in treadmill studies on AFT road shoes but different from preliminary findings for AFT spikes. Treadmill studies on road shoes consistently report that runners take ~1% longer steps in AFT shoes than in traditional shoes [9,14,21,35]. Since running speed is fixed between conditions in such studies, these longer steps are accompanied by reductions in step frequency. In a case study on overground, track running, Russo et al., [36] evaluated biomechanics during 80m trials at 95% of maximal speed comparing AFT spikes with a carbon-fiber plate and a modern midsole foam against traditional spikes. The athlete ran at the same speed and did not substantially change step length or step frequency, even though contact time was longer in the AFT spikes. The fixed speed and single subject nature of Russo et al., [36] might explain the differences with our study.

4.5 Limitations and Future Directions

While our study shows that our novel interval-based approach to quantify potential benefits of middle-distance AFT spikes is valid and reliable at the group level, we would caution against using the results of a single session to label an individual as a responder/non-responder to AFT spikes, or to compare differences in responses between individuals. Specific to that purpose, we found that at least two days are necessary to be able to evaluate individual responses to AFT spikes.

Additionally, there are limitations with the implementation of this protocol. In this current study, interval times were recorded by hand, which is subject to human error. Step lengths were calculated as the average of the 200m run, based on the average step frequency and the interval time. Future studies could place IMUs at the feet, which allows to accurately calculate stride length at a stride-by-stride level [37]. Evaluating between-spike differences in stride length at a stride-by-stride level could provide additional information about how benefits of AFT spikes potentially vary from the acceleration phase to the top speed phase.

For researchers looking to implement this protocol in future studies, great care must be taken to limit the bias of the researchers and participants. Participants should be blinded to their shoe condition. During collections in this present study, participants were typically unaware of how many different spike conditions they were running in, what spikes they were currently wearing, or that they were wearing other spikes than they thought. However, with AFT spikes becoming more common, especially when comparing commercially available spikes, some participants might be able to feel / recognize differences in midsole cushioning or longitudinal bending stiffness between spikes, which could bias their performances in our effort-based protocol.

Moreover, caution should be taken when interpreting the results on the benefits of AFT spikes in Experiment 4. On average the male participants in this experiment ran their intervals at ~6.4m/s, which is equivalent to a 2:05min 800m, or 3:54min 1500m. To evaluate the benefits at elite or even world record level, this experiment should be repeated with elite runners. Still, our findings are more informative than running economy measurements for running at 16km/h (4.44m/s; equivalent to a 3:00min 800m). Elite and/or female runners might have different plantar flexor strength or body mass than our male participants, factors are likely to affect responses to AFT [38,39]. At the start of our study, we only had spikes available in the men's sample size US9 (except for Experiment 1), but follow-up studies should include more female participants.

5 Conclusions

To overcome the challenges in quantifying the potential benefits of advanced footwear technology spikes related to the physiology of middle-distance running intensity, we introduced a novel interval-based approach. In a series of experiments, we showed the validity and reliability of our method and then used it to compare commercially available spikes. Our reliability experiment showed that group-level results were consistent across days, but that at least two separate sessions are needed to evaluate individual responses. Participants ran ~2% faster and with ~2% longer steps in two commercially available AFT spike models than in traditional spikes, while speed and step length was similar between a third AFT spike model and the traditional spikes.

Acknowledgements

We would like to thank the athletes of Athletics Canada and the UNB Reds Varsity Track and Cross Country team, and of the running communities in Amherst, Boston, San Antonio, Castilla La-Mancha and Madrid for their participation in our study. We would like to thank Rodger Kram and Jose van der Veen for energetic(s) discussions to spark the ideas presented in this paper. We would also like to thank Dan Feeney for guidance regarding our statistical analysis. Lastly, we would like to thank our undergraduate researchers, Elizabeth Hoenes, Caleb Zylka, Violeta Muñoz de la Cruz and Robert Carew for assisting with data collections and recruitment. All spikes were provided by PUMA SE.

References

1. Taylor B. What breaking the 4-minute mile taught us about the limits of conventional thinking. *Harv Bus Rev* 2018.
2. Martin Fritz Huber. The four-minute mile is still worth celebrating. *Outside* 2023;
3. Healey L, Bertschy M, Kipp S, Hoogkamer W. Can we quantify the benefits of "super spikes" in track running? *Sports Med* 2022;**52**:1211–8. https://doi.org/10.1007/s40279-022-01657-4.
4. Senefeld JW, Haischer MH, Jones AM, Wiggins CC, Beilfuss R, Joyner MJ, et al. Technological advances in elite marathon performance. *J Appl Physiol* 2021;**130**:2002–8. https://doi.org/10.1152/japplphysiol.00002.2021.
5. Rodrigo-Carranza V, González-Mohíno F, Santos-Concejero J, González-Ravé JM. Impact of advanced footwear technology on elite men's in the evolution of road race performance. *J Sports Sci* 2023:1–8. https://doi.org/10.1080/02640414.2023.2183103.
6. Rodrigo-Carranza V, González-Mohíno F, Santos del Cerro J, Santos-Concejero J, González-Ravé JM. Influence of advanced shoe technology on the top 100 annual performances in men's marathon from 2015 to 2019. *Sci Rep* 2021;**11**:1–9. https://doi.org/https://doi.org/10.1038/s41598-021-01807-0.
7. Willwacher S, Mai P, Helwig J, Hipper M, Utku B, Robbin J. Does advanced footwear technology improve track and road racing performance? An explorative analysis based on the 100 best yearly performances in the world between 2010 and 2022: The effects of super-footwear on world-class track and road racing performances. *sportrxiv* 2024;**Preprint**. https://doi.org/https://doi.org/10.51224/SRXIV.297.
8. Oehlert GM, Jones EJ, Burns GT, Joubert DP. Comparative effects of advanced footwear technology on running economy in track spikes and racing shoes: 161. *Med Sci Sports Exerc* 2023;**55**:41–2.
9. Barnes KR, Kilding AE. A randomized crossover study investigating the running economy of highly-trained male and female distance runners in marathon racing shoes versus track spikes. *Sports Med* 2019;**49**:331–42. https://doi.org/10.1007/s40279-018-1012-3.
10. Hoogkamer W, Kipp S, Spiering BA, Kram R. Altered running economy directly translates to altered distance-running performance. *Med Sci Sports Exerc* 2016;**48**:2175–80. https://doi.org/10.1249/MSS.0000000000001012.
11. Kipp S, Kram R, Hoogkamer W. Extrapolating metabolic savings in running: Implications for performance predictions. *Front Physiol* 2019;**10**:1–8. https://doi.org/10.3389/fphys.2019.00079.
12. Lucia A, Esteve-Lanao J, Olivan J, Gomez-Gallego F, San Juan AF, Santiago C, et al. Physiological characteristics of the best Eritrean runners—exceptional running economy. *Appl phy, nutri, and metab* 2006;**31**:530–40.
13. Joyner, Hunter SK, Lucia A, Jones AM. Physiology and fast marathons. *J Appl Physiol* 2020;**128**:1065–8. https://doi.org/10.1152/japplphysiol.00793.2019.

14. Hoogkamer W, Kipp S, Frank JH, Farina EM, Luo G, Kram R. A comparison of the energetic cost of running in marathon racing shoes. *Sports Med* 2018;**48**:1009–19. https://doi.org/10.1007/s40279-017-0811-2.
15. Daniels J. Daniels' running formula. Human Kinetics, 2013.
16. Haugen T, Sandbakk Ø, Seiler S, Tønnessen E. The training characteristics of world-class distance runners: an integration of scientific literature and results-proven practice. *Sports Med-Open* 2022;**8**:1–18.
17. McKay AKA, Stellingwerff T, Smith ES, Martin DT, Mujika I, Goosey-Tolfrey VL, et al. Defining training and performance caliber: a participant classification framework. *Int J Sports Physiol Perform* 2022;**17**:317–31. https://doi.org/10.1123/IJSPP.2021-0451.
18. Spiriev B, Spiriev A. IAAF scoring tables of athletics. International Association of Athletics Federations Monaco, 2011.
19. Frederick EC. Physiological and ergonomics factors in running shoe design. *Appl Ergon* 1984;**15**:281–7.
20. Franz JR, Wierzbinski CM, Kram R. Metabolic cost of running barefoot versus shod: Is lighter better? *Med Sci Sports Exerc* 2012;**44**:1519–25. https://doi.org/10.1249/MSS.0b013e3182514a88.
21. Hoogkamer W, Kipp S, Kram R. The biomechanics of competitive male runners in three marathon racing shoes: a randomized crossover study. *Sports Med* 2019;**49**:133–43. https://doi.org/10.1007/s40279-018-1024-z.
22. Day EM, Alcantara RS, McGeehan MA, Grabowski AM, Hahn ME. Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running. *J Biomech* 2021;**119**:110323.
23. Wilkinson RD, Mazzo MR, Feeney DF. Rethinking the statistical analysis of neuromechanical data. *Exerc Sport Sci Rev* 2023;**51**:43–50.
24. Cohen J. Statistical Power Analysis. *Curr Dir Psychol Sci* 1992;**1**:98–101. https://doi.org/10.1111/1467-8721.ep10768783.
25. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:14065823* 2014.
26. Lenth R, Singmann H, Love J, Buerkner P, Herve ME. Estimated marginal means, aka least-squares means R package version 1.8. 4-1. 2019.
27. Rodrigo-Carranza V, González-Mohíno F, Santos-Concejero J, González-Ravé JM. Influence of shoe mass on performance and running economy in trained runners. *Front Physiol* 2020;**11**:1–8. https://doi.org/10.3389/fphys.2020.573660.
28. Chollet M, Michelet S, Horvais N, Pavailler S, Giandolini M. Individual physiological responses to changes in shoe bending stiffness : a cluster analysis study on 96 runners. *Eur J Appl Physiol* 2022;**123**:189–177. https://doi.org/10.1007/s00421-022-05060-9.
29. Knopp M, Muñiz-Pardos B, Wackerhage H, Schönfelder M, Guppy F, Pitsiladis Y, et al. Variability in running economy of kenyan world-class and european amateur male runners with advanced footwear running technology:

experimental and meta-analysis results. *Sports Med* 2023. https://doi.org/10.1007/s40279-023-01816-1.

30. Rodrigo-Carranza V, Hoogkamer W, Salinero JJ, Rodríguez-Barbero S, González-Ravé JM, González-Mohíno F. Influence of running shoe longitudinal bending stiffness on running economy and performance in trained and national level runners. *Med Sci Sports Exerc* 2023.
31. Rodrigo-Carranza V, Hoogkamer W, González-Ravé JM, González-Mohíno F. Relationship between advanced footwear technology longitudinal bending stiffness and energy cost of running . *Med Sci Sports Exerc* 2024;**In press**.
32. Barrons Z, Rodrigo-Carranza V, Bertschy M, Hoogkamer W. The Fallacy of Single Trials: The Need for Multiple Trials in Assessing Running Economy Responses in Advanced Footwear Technology. *Sport Med* 2024;**Under review**.
33. Wood DW. An experimental test of the Puma model number 296 brush spike shoe and the Mexico Puma model number 295 standard four spike shoe as to their effect on sprint running. [Ohio]: Ohio University, 1972.
34. Hébert-Losier K, Finlayson SJ, Driller MW, Dubois B, Esculier JF, Beaven CM. Metabolic and performance responses of male runners wearing 3 types of footwear: Nike Vaporfly 4%, Saucony Endorphin racing flats, and their own shoes. *J Sport Health Sci* 2021;**00**:1–10. https://doi.org/10.1016/j.jshs.2020.11.012.
35. Hunter I, McLeod A, Valentine D, Low T, Ward J, Hager R. Running economy, mechanics, and marathon racing shoes. *J Sports Sci* 2019;**37**:2367–73. https://doi.org/10.1080/02640414.2019.1633837.
36. Russo L, Montagnani E, Buttari D, Ardigò LP, Melenco I, Larion A, et al. Track running shoes: a case report of the transition from classical spikes to “super spikes” in track running. *Appl Sci* 2022;**12**:10195.
37. Aristizabal-Pla G. Assessing the validity of the zero-velocity update method for sprinting speeds. [Amherst]: University of Massachusetts Amherst, 2022.
38. Willwacher S, König M, Braunstein B, Goldmann JP, Brüggemann GP. The gearing function of running shoe longitudinal bending stiffness. *Gait Post* 2014;**40**:386–90. https://doi.org/10.1016/j.gaitpost.2014.05.005.
39. Ortega JA, Healey LA, Swinnen W, Hoogkamer W. Energetics and biomechanics of running footwear with increased longitudinal bending stiffness: a narrative review. *Sports Med* 2021;**51**:873–94. https://doi.org/10.1007/s40279-020-01406-5.

Figures

Fig. 1 Spike conditions for each experiment. From top to bottom, Experiment 1: TRAD (PUMA EvoSpeed Distance 9), TRAD+ (PUMA EvoSpeed Distance 9 + 200g); Experiment 2: TRAD, AFTP1 (PUMA AFT middle-distance prototype spike); Experiment 3: TRAD, AFTP2 (PUMA AFT long distance prototype spike), AFTP3 (PUMA AFT middle distance prototype spike); Experiment 4: TRAD, AFTC1 (PUMA evoSPEED Long Distance Nitro Elite+), AFTC2 (Competitor brand AFT spike), AFTC3 (Competitor brand AFT spike).

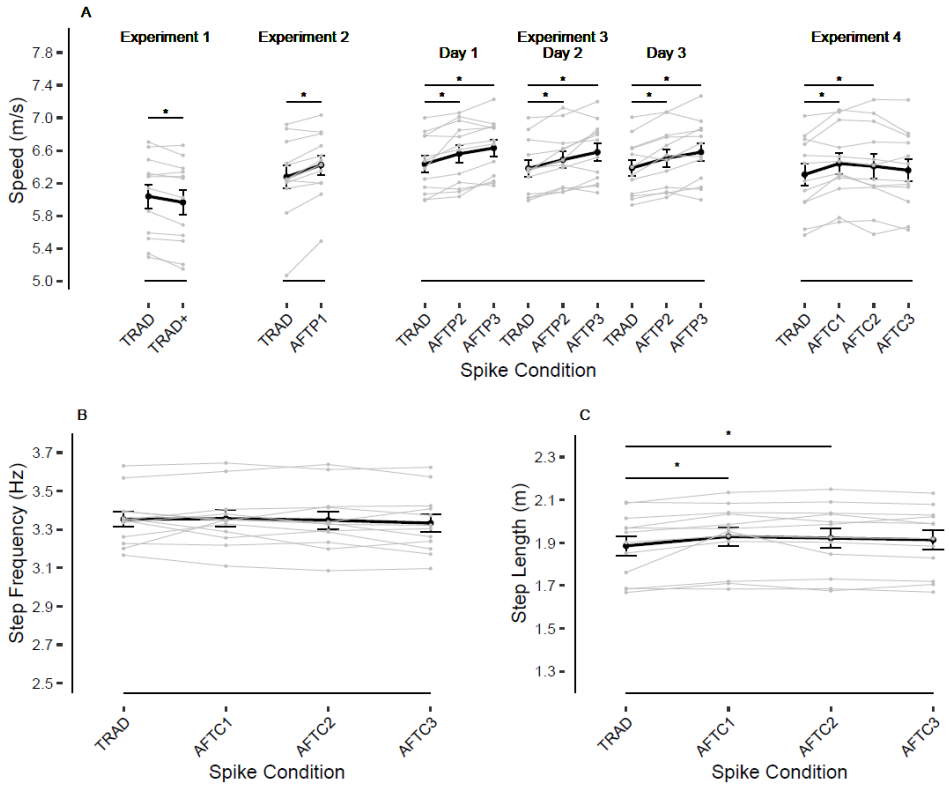

Fig. 2 Spike conditions significantly affect running speed and step length, but not step frequency. Panel **A** shows average running speeds in different spike conditions, for Experiment 1, 2, 3 and 4, from left to right. Panel **B** shows significantly longer step lengths for AFTC1 and AFTC2 than TRAD. Panel **C** shows step frequency was similar across spike conditions. Grey lines represent individual data, black lines represent group mean values, error bars represent standard error of the mean. * indicates significant difference between spike conditions.

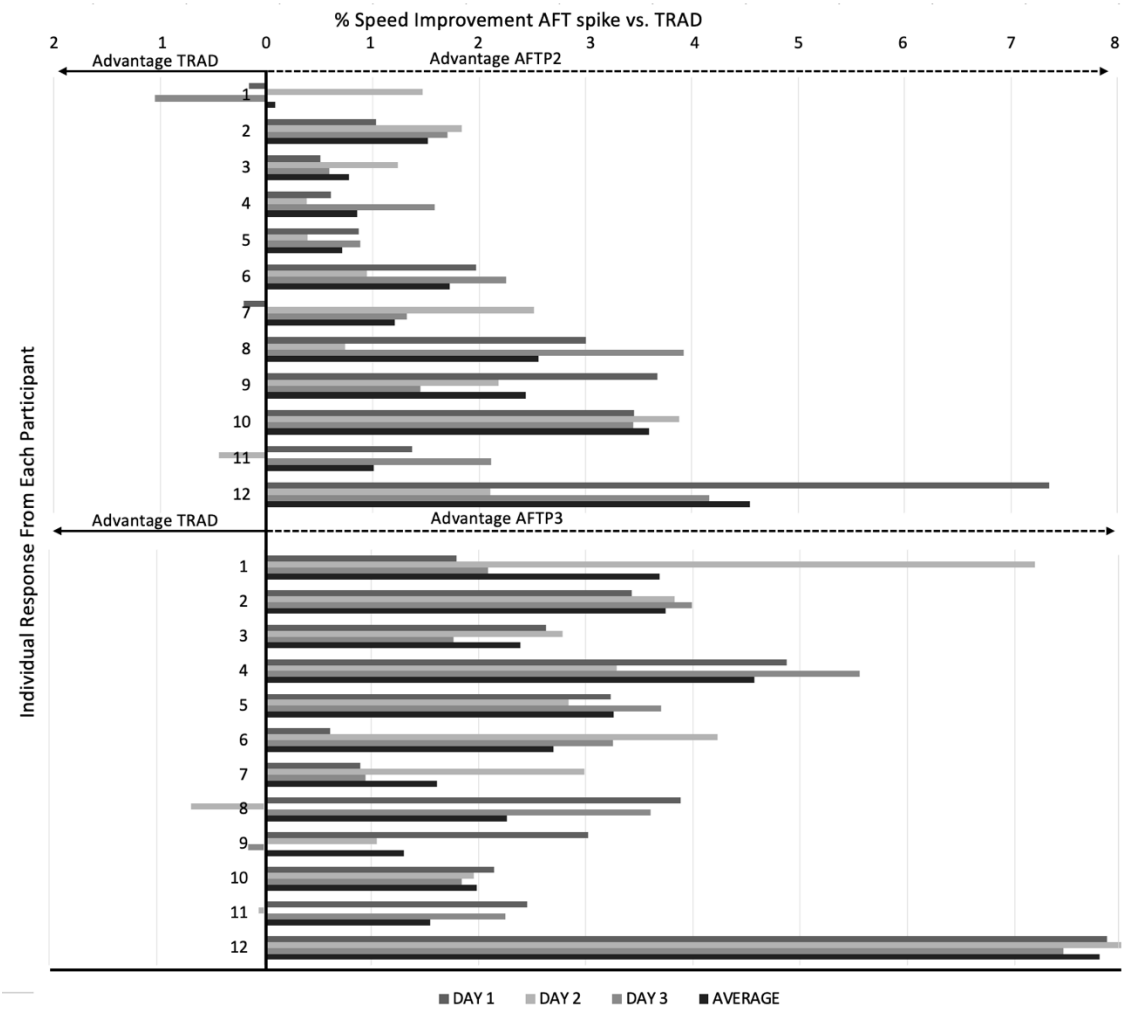

Fig. 3 Individual response of each participant in prototype AFT spikes vs. traditional spikes (TRAD) in Experiment 3 for each of the three days and the 3-day average.

Tables

Table 1. Participant characteristics for each experiment (mean ± standard deviation).

Experiment	Participants (men, women)	Age (years)	Mass (kg)	Height (m)	Location
1	8, 4	20.3 ± 2.3	66.2 ± 10.3	1.76 ± 0.09	US
2	12, 0	21.3 ± 1.8	62.8 ± 2.8	1.75 ± 0.04	US, Canada
3	12, 0	27.0 ± 5.7	64.8 ± 3.3	1.76 ± 0.06	Spain
4	12, 0	22.0 ± 2.7	69.1 ± 8.3	1.81 ± 0.07	Canada

Table 2: Characteristics of experimental conditions used in each Experiment.

Spike	Sizes	Mass (g)	Midsole thickness (mm)	Experiments used
TRAD	5.5, 7, 9, 11	131, 149, 158, 170	17	1,2,3,4
TRAD+	5.5, 7, 9, 11	331, 341, 355, 369	17	1
AFTP1	9	179	19	2
AFTP2	9	151	18	3
AFTP3	9	149	19	3
AFTC1	9	154	18	4
AFTC2	9	140	12	4
AFTC3	9	163	17	4

Table 3: Magnitude and significance of between spike differences were consistent across days.

Comparison		Day 1	Day 2	Day 3	3-Day average
TRAD vs. AFTP2	% change	2.02	1.72	1.74	1.87
	effect size	0.36	0.32	0.31	0.34
	p value	0.003	0.004	0.0003	<0.0001
TRAD vs. AFTP3	% change	3.11	3.13	2.93	3.12
	effect size	0.57	0.56	0.51	0.57
	p value	<0.0001	<0.0001	<0.0001	<0.0001
AFTP2 vs. AFTP3	% change	1.06	1.38	1.23	1.22
	effect size	0.19	0.25	0.21	0.19
	p value	0.124	0.015	0.035	0.0004