2017 American Control Conference
Sheraton Seattle Hotel
May 24–26, 2017, Seattle, USA

System Identification of *Just Walk*: A Behavioral mHealth Intervention for Promoting Physical Activity

Mohammad T. Freigoun[1], César A. Martín[1,2], Alicia B. Magann[1], Daniel E. Rivera[1], Sayali S. Phatak[3], Elizabeth V. Korinek[3], and Eric B. Hekler[3]

***Abstract*— There is significant evidence to show that physical activity reduces the risk of many chronic diseases. With the rise of mobile health (mHealth) technologies, one promising approach is to design interventions that are responsive to an individual's changing needs. This is the overarching goal of *Just Walk*, an intensively adaptive physical activity intervention that has been designed on the basis of system identification and control engineering principles. Features of this intervention include the use of multisine signals as pseudo-random inputs for providing daily step goals and reward targets for participants, and an unconventional ARX estimation-validation procedure applied to judiciously-selected data segments that seeks to balance predictive ability over validation data segments with overall goodness of fit. Analysis of the estimated models provides important clues to individual participant characteristics that influence physical activity. The insights gained from black-box modeling are critical to building semi-physical models based on a dynamic extension of Social Cognitive Theory.**

I. INTRODUCTION

One recent emerging application of system identification and control theory is the design of optimized interventions in health behavior. Strong evidence shows that physical activity (PA) reduces chronic disease risk [1], [2], [3]. National guidelines suggest 30-60 minutes of moderate PA per day, often in the form of walking [4]. Research furthermore demonstrates 20-30% reduced risk of breast cancer when guidelines are met [5]. While the benefits of PA are well known, a large segment of the population does not meet these guidelines [6]. Thus, an important scientific problem is not *if* PA influences health, but rather *how* to promote favorable and sustained PA levels, particularly among sedentary middle-aged adults. Strategies to fit regular PA into one's life are idiosyncratic and dynamic; system identification results that support this conclusion follow in this paper. Not only is the behavior idiosyncratic, but so are the times and places when individuals fit PA into their daily routine. Mobile health (mHealth) technologies offer a cost-effective, convenient, and scalable platform for delivering information-intensive and perpetually adaptive PA interventions.

[1]Control Systems Engineering Laboratory, School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ 85287 USA m.freigoun@asu.edu, cmartinm@asu.edu, alicia.magann@asu.edu, daniel.rivera@asu.edu

[2]Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador cmartin@espol.edu.ec

[3]School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ 85004, USA sayali.phatak@asu.edu, elizabeth.korinek@asu.edu, ehekler@asu.edu

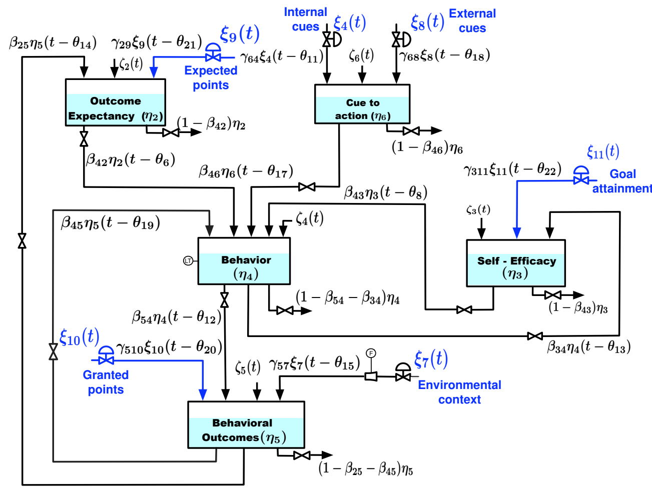

Fig. 1. Dynamical fluid analogy model of Social Cognitive Theory [9].

In designing adaptive interventions [7], a key consideration is the ability to estimate *personalized* behavior models that identify both individual-invariant and individual-variant dynamics. Parsimonious modeling, guided by *a priori* knowledge, is therefore crucial to accomplishing this. Using control systems engineering principles, behavior change theory can be utilized to develop models and decision frameworks for interventions that promote PA among sedentary individuals. One example is Social Cognitive Theory (SCT), proposed by Bandura [8], which is among the leading theories of behavior change. The work of Martín *et al.* [9] established a dynamical systems fluid analogy model that captures key SCT concepts. Fig. 1 represents a simplified fluid analogy dynamical system model of SCT [9]. SCT includes an extended list of potential constructs associated with predicting complex behavior dynamics, and whose effect needs to be accounted for during modeling. Consider, for example, the *Environmental Context* construct in Fig. 1; this can include weather, busyness, stress, weekday, mood, and several other known and unknown variables.

Following an experimental design methodology based on system identification principles [10], a unique single-subject intervention study, *Just Walk*, was performed. Section II introduces the *Just Walk* intervention and experiment execution, followed by a brief description on the input signal design approach using orthogonal multisine excitations in Section III. Section IV discusses the use of an unconventional black-box approach that provides key insights into the

978-1-5090-5992-8/$31.00 ©2017 AACC

Fig. 2. Screenshot of the *Just Walk* app.

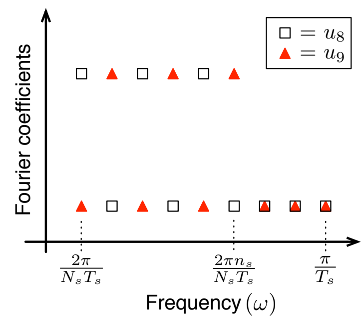

Fig. 3. Conceptual representation of a "zippered" spectra design for $n_u = 2$ design inputs, and $n_s = 6$ harmonic frequencies.

dynamics of the intervention participants, and will ultimately be instrumental in accomplishing semi-physical identification of SCT models (Fig. 1). Section V outlines a number of important conclusions and future directions on input signal design, modeling, identification, and intervention design.

II. DESCRIPTION OF *JUST WALK* INTERVENTION

Just Walk was developed as an adaptive walking intervention app for sedentary, overweight adults. It was designed primarily as a tool to generate individualized computational models for understanding PA behavior via system identification. The intervention system included a front-end Android app, *Just Walk* (Fig. 2), a backend server, and an activity tracker (Fitbit Zip) to objectively measure PA and automatically sync with the smartphone application. Participants were recruited nationally to partake in a walking intervention and receive daily step goals via the *Just Walk* app, and daily announced points were granted if the goals were achieved that day; granted points were converted into Amazon gift cards after a certain threshold was reached. Participants were also required to complete a series of daily morning and evening ecological momentary assessment (EMA; [11]) measures (e.g. confidence in achieving goal, predicted business for that day, previous night's sleep quality, etc.) for the entire duration of the study.

The study duration was 14 weeks, including an initial two-week baseline period in which no step goals were delivered. Each participant's step goals were then based on their median daily step value as calculated from the 14-day baseline period. The step goals were designed to establish a mechanism for individualizing the definition of an "ambitious, but doable" step range. All PA data were collected from the Fitbit Zip (provided to participants as a part of the study) and stored both locally and in Fitabase (Small Steps Labs, San Diego, CA, USA). Participants were generally healthy, inactive, 40-65 years old, with a body mass index (BMI) of 25-45 kg/m^2, who currently owned an Android phone capable of connecting to a Fitbit Zip via Bluetooth 4.0, and were willing to engage with the mHealth intervention for 14 weeks.

III. INPUT SIGNAL DESIGN OF *JUST WALK*

The input signal design procedure utilized in the *Just Walk* study [12] was designed using deterministic yet "pseudo-random" signals that are orthogonal in the frequency domain. The procedure is described in detail in [10]. In *Just Walk*, *Goals* establish the desired behavior in a quantitative form, while *Expected Points* are the daily available points announced each morning that are granted upon goal achievement. *Goals* and *Expected Points* are two manipulated input signals u_n generated from a multisine signal,

$$u_n(k) = \lambda_n \sum_{j=1}^{N_s/2} \sqrt{2\alpha_{[n,j]}} \cos(\omega_j k T_s + \phi_{[n,j]}) \\ \omega_j = \frac{2\pi j}{N_s T_s}, \quad k = 1, \ldots, N_s \tag{1}$$

where λ_n is the scaling factor, N_s is the number of samples per period, T_s is the sampling time. For the j^{th} harmonic of the signal each variable has the following meaning: $\alpha_{[n,j]}$ is a factor used to specify the relative power of the harmonic, ω_j is the frequency, and $\phi_{[n,j]}$ is the phase. To obtain independent transfer function and uncertainty estimates, factors $\alpha_{[n,j]}$ are chosen to excite input signals orthogonally in frequency. Two signals are orthogonal if a nonzero Fourier coefficient at a specific frequency in one signal implies a zero-valued Fourier coefficient at the same frequency for the other; this is called a "zippered" spectra design, an idea introduced in [13]. A conceptual representation of the "zippered" design is presented in Fig. 3. For n_u design inputs and n_s independently excited sinusoids the Fourier coefficients are specified as

$$\alpha_{[n,j]} = \begin{cases} 1 & \text{if } j = n_u(i-1) + (n-7) \\ & \text{for } i = 1, 2, \ldots, n_s \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Using the ω_j frequencies defined in (1) and the Nyquist-Shannon sampling theorem, the following bound for N_s is defined:

$$N_s \geq 2n_s \tag{3}$$

If $n_s = 6$ excited sinusoids are selected for the $n_u = 2$ design inputs, then from applying (3), $N_s = 16$ days (selected) is a feasible option. Phases $\phi_{[n,j]}$ are selected to minimize the crest factor of the signal using the approach proposed by Guillaume *et al.* [14].

In applying this design methodology for *Just Walk*, amplitudes for input signals (u_8 and u_9 in Fig. 4) were chosen relying on experiences from previous studies [15], [16] designed to obtain an expected profile of PA. The maximum number of step goals was selected as a factor of the initial baseline level of PA. For most cases in this experimental design, this factor was equal to 2; however, it was varied if the actual baseline step level of individuals was too high or low. Specifically, if participant's baseline median steps were below 3,000, then the range for the goals was between 1 and 2.5 of their baseline median steps, to increase the likelihood of "ambitious" goals. If baseline median steps were greater than 7,500 steps, then the range was set between 1 and 1.75 (to reduce the likelihood of overly ambitious goals, such as 15,000 steps in one day). In addition to the two manipulated input channels, a large set of disturbances were also measured using mHealth technologies. Overall experimental duration beyond the baseline varied between five to six cycles for each participant. A time series plot for a representative participant that depicts the behavior and seven inputs is shown in Fig. 4.

IV. ARX MODEL ESTIMATION & VALIDATION

In this section, black-box modeling strategies used for *Just Walk* are outlined, and results from fitting Auto Regressive with eXogenous input (ARX) parametric models [17] are presented. As noted, identifying optimal ARX models (decisions on model inputs and model order) will play a pivotal role in ultimately identifying *personalized* semi-physical (grey-box) models that are informed by well-established behavior theories [9]. Prior to ARX estimation, standard nonparametric modeling tasks such as correlation analysis have been informative. Because the *Just Walk* study included a wide array of input/output measurements, results from input-output and input-input correlation analyses have been useful [18], [12], [19]; for reasons of brevity these are not included in this paper. Incorporating all measured disturbances for estimating an SCT behavior model (particularly the *Environmental Context* construct in Fig. 1) can be computationally demanding, may pose *identifiability* challenges, and will require large informative datasets that are typically difficult to gather from a practical standpoint in research involving human subjects.

Preprocessed data are fitted to an ARX model structure, ARX-[n_a,n_{b_1},…,$n_{b_{n_u}}$,n_{k_1},…,$n_{k_{n_u}}$] which can be expressed in the following concise form:

$$y(t) + \sum_{l=1}^{n_a} a_l y(t-l) = \sum_{j=1}^{n_u} \sum_{i=0}^{n_{b_j}-1} b_{(i+1)(j)} u_j \left(t - n_{k_j} - i\right) + e(t) \tag{4}$$

where $y(t)$ is the measured output (e.g., steps/day), $u_j(t)$ is the measured input j, $e(t)$ is the prediction error, all measured/estimated at day t. The ARX model in (4) is estimated by using regression. ARX parameter estimation constitutes a linear least-squares regression problem [17] and has attractive statistical properties such as consistency. Fig. 4 illustrates an example contrasting the difference between actual output measurements and the prediction from a 7-input ARX model with the structure in (4); a detailed discussion of black-box modeling strategies used in this work follows. To quantify model fits, the normalized root mean square error (NRMSE) fit index is used

$$\text{model fit } (\%) = 100 \times \left(1 - \frac{||y(k) - \hat{y}(k)||_2}{||y(k) - \bar{y}||_2}\right) \tag{5}$$

$y(k)$ is the measured output, $\hat{y}(k)$ is the simulated output, $\bar{y}$ is the mean of all measured $y(k)$ values, and $||\cdot||_2$ indicates a vector l_2-norm.

A. Data Pre-Processing and Model Structure Considerations

Data pre-processing tasks include interpolation (to account for missing data), mean subtraction, and shifting *Actual Steps* and *Granted Points* by one sample to reflect temporal precedence. Model structure selection decisions consist of determining, for each participant, the input signals to be included, and corresponding ARX model orders for the output and each input, in accordance with (4). Taking advantage of the computational simplicity associated with ARX modeling, the approach taken here is to exhaustively examine a range of model orders, and use model validation procedures to determine the most suitable structure. For this case study, ARX model order ranges for n_a and n_b from 1 to 3 (i.e., $\max(n_a) = 3$, and $\max(n_{b_j}) = 3 \quad \forall j = 1, \ldots, n_u$) seemed reasonable. *A priori* knowledge of the SCT fluid analogy model developed in [9] implies that very high order models should not be necessary to characterize these behavior-change dynamics. From inspecting the intervention data, it was reasonable to assume a basic unit input delay (i.e., $n_{k_j} = 1 \quad \forall j$). Finally, the absence of drifts in the data leads to assume stationary (though potentially time-varying) noise characteristics over the course of the intervention period.

In determining the inputs to be considered, the approach is to start with a basic 3-input model consisting of *Goals* (u_8), *Expected Points* (u_9), and *Granted Points* (u_{10}) and then add 4 additional measured inputs (*Predicted Busyness*, *Predicted Stress*, *Predicted Typical*, and *Weekday-Weekend*) to this basic model. All possible combinations of these inputs are estimated. Model validation following estimation ultimately determines which of these inputs are most important in describing individual behavior. Nonetheless, in the pre-processing stage, correlation analysis can be used to determine inputs that may be significantly crosscorrelated with each other or to identify inputs that appear to have no significant effect on the output. In both scenarios, the number of inputs that needs to be considered in the parameter estimation procedure can be reduced, ultimately leading to parsimonious models that can be generated with less effort.

B. Model Parameter Estimation and Validation

Model estimation and concomitant validation with the *Just Walk* intervention data is now considered. As mentioned earlier, first, a core 3-input model was estimated and evaluated, followed by the addition and combination of 4 more inputs, leading to estimation of all possible combinations of these

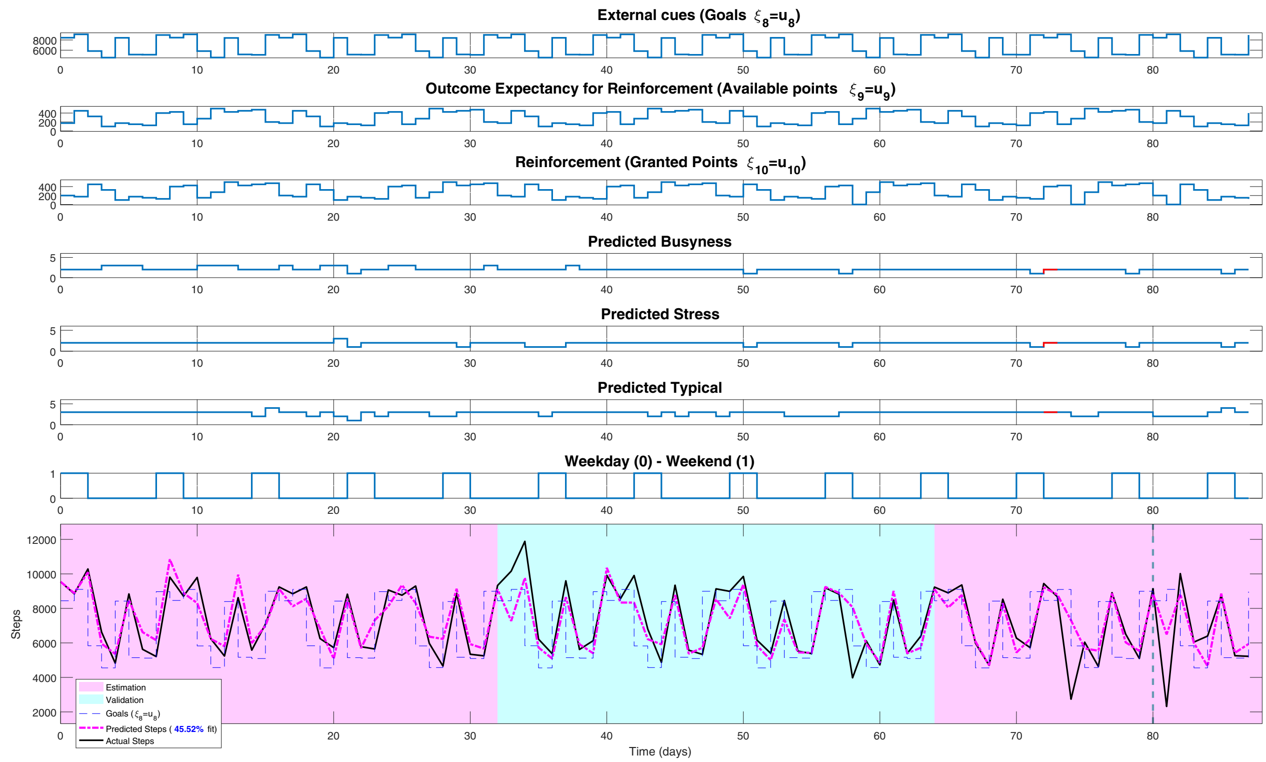

Fig. 4. Time series plot showing seven selected input sequences (manipulated inputs & measured disturbances), predicted behavior (from an ARX black-box model), actual behavior, model overall fit, and estimation & validation cycles (1st, 2nd, and 5th for estimation; 3rd and 4th for validation) for a selected *Just Walk* participant.

additional inputs. At an individual level, the full dataset was segmented into informative 16-day cycles for model estimation/validation. The cycle length was defined by the multisine input signal described in Section III.

Cross-validation (the process of evaluating model fit over data not used for estimation) represents one of the most valuable aspects of system identification [20]. The conventional approach in system identification is to assign a certain percentage of data for estimation, followed by validation (e.g., 50% estimation, 50% validation). Such an approach assumes that the noise characteristics of the problem remain unchanged during the course of the intervention. However, it is reasonable to expect that noise and disturbance characteristics will vary over long-duration interventions such as *Just Walk*. In the analysis, each data cycle was assigned to either estimation or validation; all combinations of data cycles involving at least two cycles for validation were generated and evaluated.

Table I summarizes results of this procedure for a 4-input model (*Goals*, *Expected Points*, *Granted Points* and *Predicted Busyness*) of a selected participant. The fit index from Equation (5) was calculated for each cycle and averaged for estimation and validation data, respectively. All data cycle combinations that feature at least two cycles for validation or estimation (twenty candidate ARX models) were evaluated. For each of these combinations of estimation and validation cycles (corresponding to a specific row in Table I), ARX orders were determined from an exhaustive search routine that selects a stable ARX model with highest predictive ability (based on the maximum average validation fit). This step provides a safeguard against overparametrization. The final chosen model should reflect, in addition to a good fit to validation data, a good fit for the entire data set (consisting of both estimation and validation cycles). This suggests that the final model choice should correspond to the model that yields highest overall fit (the "Overall NRMSE Fit" column in Table I). Incorporating the overall fit criterion with the fit to cross-validation data balances good prediction with model accuracy over the entire data set. Note that using this analysis, the best results for the specific participant occur in the model resulting from row 18 (cycles 1, 2, and 5 for estimation; 3 and 4 for validation) with an overall NRMSE index at 46.03% for a model with structure $n_a = 2$, $n_{b_1} = 3$, $n_{b_2} = 1$, $n_{b_3} = 2$, and $n_{b_4} = 3$. This model performs close to the model with best fit over the validation data (average validation fit of 56.63% for row 18 vs 60.65% in row 15); however, the model with the best fit to validation data does not exhibit the best fit to data overall (38.81% in lieu of 46.03%).

C. Overall Fit Analysis and Assessment of Individual Participant Characteristics

Similar analyses to those presented in Table I can be performed with additional inputs, for all possible combinations. For example, for a total of 7 inputs, 16 different

TABLE I

INTERMEDIATE RESULTS FOR A 4-INPUT ARX MODEL OF A SELECTED PARTICIPANT FROM *Just Walk*

E*	V*	NRMSE Fit (%)					Average Estimation NRMSE Fit (%)	Average NRMSE Validation Fit (%)	Overall NRMSE Fit (%)	ARX Order (4-input) $[na_1,nb_1,nb_2,nb_3,nb_4]$
		Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5				
[1,2]	[3,4,5]	77.40%	85.44%	79.27%	27.68%	13.70%	81.42%	40.22%	40.11%	[1,1,1,1,3]
[1,3]	[2,4,5]	77.39%	82.25%	81.30%	26.88%	15.36%	79.35%	41.50%	40.60%	[1,2,1,1,3]
[1,4]	[2,3,5]	64.82%	71.25%	67.27%	45.89%	21.04%	55.36%	53.19%	42.29%	[1,3,1,1,1]
[1,5]	[2,3,4]	61.36%	59.51%	60.96%	40.14%	24.47%	42.92%	53.54%	37.40%	[1,1,1,3,1]
[2,3]	[1,4,5]	70.46%	90.25%	84.15%	25.00%	11.19%	87.20%	35.55%	37.70%	[3,3,1,2,3]
[2,4]	[1,3,5]	49.06%	71.94%	67.25%	52.39%	22.98%	62.17%	46.43%	40.56%	[3,1,2,1,3]
[2,5]	[1,3,4]	54.89%	61.75%	60.36%	47.35%	23.68%	42.72%	54.20%	39.33%	[3,1,1,1,1]
[3,4]	[1,2,5]	45.97%	61.27%	69.24%	51.46%	24.02%	60.35%	43.75%	41.15%	[1,3,3,1,3]
[3,5]	[1,2,4]	63.11%	66.96%	52.29%	41.52%	19.47%	35.88%	57.20%	41.12%	[1,1,1,1,1]
[4,5]	[1,2,3]	36.37%	52.47%	50.06%	49.24%	25.88%	37.56%	46.30%	32.75%	[1,1,1,3,2]
[3,4,5]	[1,2]	53.63%	64.61%	49.26%	46.59%	19.93%	38.59%	59.12%	40.12%	[1,1,1,1,1]
[2,4,5]	[1,3]	50.12%	59.76%	59.36%	49.92%	23.64%	44.44%	54.74%	38.71%	[3,1,1,1,1]
[2,3,5]	[1,4]	58.63%	66.76%	64.91%	49.62%	27.28%	52.98%	54.13%	40.59%	[3,1,3,2,1]
[2,3,4]	[1,5]	59.43%	76.99%	70.11%	41.51%	22.32%	62.87%	40.88%	41.61%	[2,3,3,2,3]
[1,4,5]	[2,3]	57.91%	61.11%	60.18%	45.69%	24.92%	42.84%	60.65%	38.81%	[1,1,1,3,1]
[1,3,5]	[2,4]	66.34%	66.02%	67.24%	42.13%	22.57%	52.05%	54.08%	41.31%	[1,3,1,1,1]
[1,3,4]	[2,5]	68.39%	77.75%	73.46%	41.86%	18.78%	61.24%	48.27%	42.26%	[1,3,2,1,1]
[1,2,5]	[3,4]	61.85%	56.05%	68.43%	44.82%	35.02%	50.97%	56.63%	46.03%	[2,3,1,2,3]
[1,2,4]	[3,5]	71.99%	73.18%	72.36%	43.28%	20.40%	62.82%	46.38%	43.61%	[1,2,1,1,3]
[1,2,3]	[4,5]	75.95%	87.02%	80.67%	26.39%	13.36%	81.21%	19.88%	39.87%	[1,1,1,1,3]

*E≡Estimation Cycles (magenta), V≡Validation Cycles (cyan)

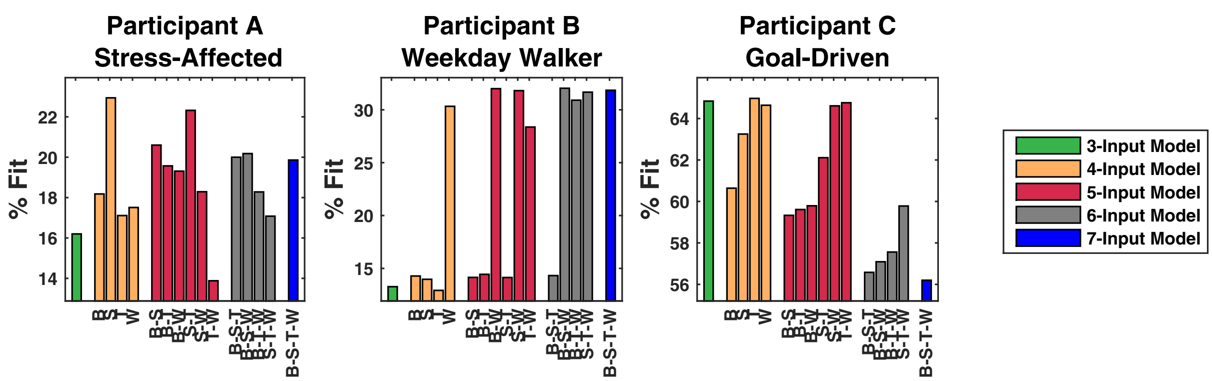

Fig. 5. Average validation % fits of individualized ARX models from black-box system identification for three individuals: Goals-Expected Points-Granted Points model; **B**: Predicted Busyness; **S**: Predicted Stress; **T**: Predicted Typical; **W**: Weekday-Weekend.

input models can be generated for each participant (since *Goals* (u_8), *Expected Points* (u_9), and *Granted Points* (u_{10}) are always grouped). Evaluating these 16 input combinations allows us to draw conclusions on participant characteristics that resulted from the intervention. Fig. 5 depicts model validation % fit results from three different participants from *Just Walk*. The Y-axis indicates the % fit of the 3, 4, 5, 6, and 7-input models, and the X-axis corresponds to the psychosocial measures (busyness, stress, weekday, typical) measured daily. Here, it is seen that Participant A's walking behavior is largely driven by stress (highest % fit seen for the stress bar in the 4-input model), Participant B's behavior is driven by whether it is a weekday or weekend, while Participant C has the highest % fit for the 3-input model, indicating that the daily step goal had the greatest impact on walking behavior. Step responses from the individual ARX models can be used to reveal more precise directionality and magnitude information; for example, from Fig. 6, one can predict that the selected participant will typically reach 80% of the desired daily step goals within the first day of goal announcement. Responsiveness to other inputs and disturbances can be determined similarly. This strategy has significant implications for personalized and adaptive behavior change interventions; if one can determine the inputs that are most meaningful for a given individual in a given context, it is possible then to optimize the target behavior over a specified time (hours, days, weeks, months).

V. CONCLUSIONS AND FUTURE WORK

This paper presents a system identification modeling strategy for a physical activity (e.g., walking) intervention delivered via a smartphone application. The results from this dataset represent an important accomplishment in understanding behavior-change from a data-driven perspective. Predictive and consistent black-box models are crucial for

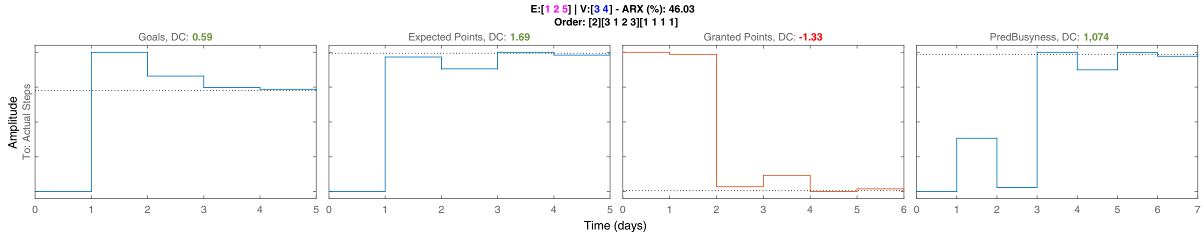

Fig. 6. Step responses of a 4-input model (including *Predicted Busyness*) for a selected participant.

validating behavioral theory (such as the SCT model). It is shown that segmenting and evaluating the data at a per-cycle level gives the most valid results to date. These types of models are necessary to effectively model behavior, which is highly complex, idiosyncratic, and dynamic in nature. In addition, drawn from the analysis of the estimated models, it is important in experimental design to consider capturing more low-frequency dynamics (e.g., using pseudo-random binary sequences), to draw more decisive conclusions on participant long-term (steady-state) responses. Finally, the enhanced identification testing monitoring procedure in [21] can be considered in future experiment design.

Guided by the analyses presented in this paper, a dynamical systems model of Social Cognitive Theory in [9] (Fig. 1) will be estimated and validated from *Just Walk* data. Additional work on *Just Walk* includes incorporating weather information, evaluate models whose inputs are intermediate signals in Fig. 1 (e.g., *self-efficacy*, *behavioral outcomes*, etc.) that can be constructed by means of aggregation of existing measurements, and modeling noticeable time-varying dynamics. To this effect, linear parameter-varying (LPV) models will be identified and validated from data (black- and grey-box models). Finally, an important future aim is to establish an LPV-HMPC framework that equips the established HMPC algorithm [22] with the ability of accommodating LPV models.

VI. ACKNOWLEDGMENTS

Support for this work has been provided by the National Science Foundation (NSF) through grant IIS-1449751. The opinions expressed in this article are the authors' own and do not necessarily reflect the views of NSF.

REFERENCES

[1] U. E. Bauer, P. A. Briss, R. A. Goodman, and B. A. Bowman, "Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA," *The Lancet*, vol. 384, no. 9937, pp. 45 – 52, 2014.

[2] N. Owen, A. Bauman, and W. Brown, "Too much sitting: a novel and important predictor of chronic disease risk?" *British Journal of Sports Medicine*, vol. 43, no. 2, pp. 81–83, dec 2008.

[3] W. L. Haskell, S. N. Blair, and J. O. Hill, "Physical activity: Health outcomes and importance for public health policy," *Preventive Medicine*, vol. 49, no. 4, pp. 280 – 282, 2009.

[4] U.S. Department of Health and Human Services. (2008) 2008 Physical Activity Guidelines for Americans. [Online]. Available: https://health.gov/paguidelines

[5] I. Thune, T. Brenn, E. Lund, and M. Gaard, "Physical activity and the risk of breast cancer," *New England Journal of Medicine*, vol. 336, no. 18, pp. 1269–1275, 1997, pMID: 9113929.

[6] R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, and M. Mcdowell, "Physical activity in the United States measured by accelerometer," *Medicine & Science in Sports & Exercise*, vol. 40, no. 1, pp. 181–188, Jan 2008.

[7] D. E. Rivera, M. D. Pew, and L. M. Collins, "Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction," *Drug and Alcohol Dependence*, vol. 88, no. Supplement 2, pp. S31 – S40, 2007, Customizing Treatment to the Patient: Adaptive Treatment Strategies.

[8] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall series in social learning theory, 1986.

[9] C. A. Martín, D. E. Rivera, W. T. Riley, E. B. Hekler, M. P. Buman, M. A. Adams, and A. C. King, "A dynamical systems model of Social Cognitive Theory," in *Proceedings of the American Control Conference*, 2014, pp. 2407–2412.

[10] C. A. Martín, D. E. Rivera, and E. B. Hekler, "Design of informative identification experiments for behavioral interventions," in *Proceedings of the 17th IFAC Symposium on System Identification*, 2015, pp. 1325–1330.

[11] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Clinical Psychology*, vol. 4, no. 1, pp. 1–32, 2008.

[12] E. B. Hekler, "Just walk study," http://justwalkstudy.weebly.com/, 2015, [Online; accessed September-23-2015].

[13] D. E. Rivera, H. Lee, H. D. Mittelmann, and M. W. Braun, "Constrained multisine input signals for plant-friendly identification of chemical process systems," *Journal of Process Control*, vol. 19, no. 4, pp. 623–635, 2009.

[14] P. Guillaume, J. Schoukens, R. Pintelon, and I. Kollar, "Crest-factor minimization using nonlinear Chebyshev approximation methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, no. 6, pp. 982–989, 1991.

[15] A. C. King, E. B. Hekler, L. A. Grieco, S. J. Winter, J. L. Sheats, M. P. Buman, B. Banerjee, T. N. Robinson, and J. Cirimele, "Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults," *PLoS ONE*, vol. 8, no. 4, p. e62613, 2013.

[16] M. A. Adams, J. F. Sallis, G. J. Norman, M. F. Hovell, E. B. Hekler, and E. Perata, "An adaptive physical activity intervention for overweight adults: A randomized controlled trial," *PloS one*, vol. 8, no. 12, p. e82901, 2013.

[17] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 1999.

[18] S. S. Phatak, E. B. Hekler, D. E. Rivera, C. A. Martín, and M. T. Freigoun, "Building a dynamical model to predict "ambitious but doable" daily step goals," in *Annual Meeting of the Society of Behavioral Medicine*, United States, Washington DC, 2016.

[19] E. V. Korinek, S. S. Phatak, C. A. Martín, M. T. Freigoun, D. E. Rivera, M. Adams, P. Klasnja, M. Buman, and E. B. Hekler, "Adaptive step goals: A longitudinal growth model of daily steps for a smartphone-based walking intervention." Submitted for Publication.

[20] L. Ljung, "From Data to Model: A Guided Tour," in *International IEE Conference on Control*, Warwick, England, Mar 1994, pp. 422–430.

[21] C. A. Martn, D. E. Rivera, and E. B. Hekler, "An enhanced identification test monitoring procedure for mimo systems relying on uncertainty estimates," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 2091–2096.

[22] C. A. Martín, D. E. Rivera, and E. B. Hekler, "A decision framework for an adaptive behavioral intervention for physical activity using hybrid model predictive control," in *Proceedings of the American Control Conference*, 2016, pp. 3576–3581.