Data-driven models of pelvic floor muscles dynamics subject to psychological and physiological stimuli

Steffi Knorn *, **Damiano Varagnolo** **, **Reinhilde Melles** ***, and **Marieke Dewitte** ***

* *Department of Engineering Sciences, Uppsala University, Sweden*
** *Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway*
*** *Department of Psychology and Neuroscience, Clinical Psychological Science, Behavioural Medicine, Maastricht Universitair Medisch Centrum, Maastricht, The Netherlands*

Abstract: This paper proposes individualized, dynamical and data-driven models to describe pelvic floor muscle responses in women undergoing vaginal dilation. Specifically, the models describe how the aggregated pressure exerted by the pelvic floor muscles of women change due to physiological and psychological stimuli. Specifically, women experienced inflation of a balloon at the vaginal introitus while watching different short movies such as with or without sexual content. The paper inspects the approximation capabilities of different model structures, such as Hammerstein-Wiener and NARX, for this specific application, and finds the specific model structures and orders that best describe the recorded measurement data. Moreover, the manuscript explores the trade-offs between individualization and averaging of models. More precisely it numerically assesses how models obtained by assuming that each individual has the same response can be used to simulate the responses of different patients. Although the current dataset is drawn from a sample of healthy volunteers, this paper is an initial step towards better understanding women's responses to vaginal dilation and sexual/nonsexual videos and facilitating individualised medical vaginal dilation treatment.

Keywords: female sexual dysfunction, black box, nonlinear models, system identification

1. INTRODUCTION

Painful experiences during sexual intercourses, also summarized with the acronym Genital pain / penetration disorders (GPPD), is a common problem and affects an estimated 30-40% of women at least once in their life (Goldstein et al., 2009, Chap. 2). The condition can be caused by physiological processes (e.g., complications after cervix cancer surgeries, vaginal radiotherapies, Mayer-Rokitansky-Küster-Hauser syndromes, male-to-female gender confirmation surgeries), psychosocial processes (e.g., traumatic sexual experiences), and combinations thereof (Goldstein et al., 2009, Chap. 3). Research in the area concludes that psychological mechanisms (e.g., anxiety, catastrophising pain, avoidance of sexual intimacy) and interpersonal factors (e.g., hostile partner responses, relationship conflicts) may maintain, prolong and exacerbate the suffering from genital pain.

Treating GPPD usually combines psychological (e.g., Cognitive Behavioral Therapies (CBTs)) and physiological therapies. The latter ones may include stretching the vaginal duct, desensitizing the vestibulum, and relaxing the pelvic floor muscles (Binik et al., 2006; Bergeron et al., 2008; Goldstein et al., 2011) by inserting opportunely sized vaginal dilators. However, these therapies are perceived as invasive, lengthy and uncomfortable. Hence, several

patients delay, avoid or stop treatment. This problem might be alleviated by individualising the vaginal dilation patterns, but to the best of our knowledge it is still unclear *how* to quantitatively perform this individualisation step and *how much* this will improve the situation. In this paper, we propose control-oriented models to facilitate designing such individual therapies.

1.1 Literature review

Before detailing our contributions, we discuss the existing literature that is related to the topic of understanding how to design vaginal dilation therapies.

First, we consider the medical literature where several studies exist, that analyse the physiological implications of stimulating the pelvic area using medical-oriented approaches. For instance, it is known that sexual arousal in women induces genital blood flow, which then leads to both vasocongestion of the vestibular bulbs Puppo (2013) and vaginal lubrication Levin (2002); Boyer (2009). Further, inducing vibrations in a "suitable range" of the inner labia and the vestibular bulbs may facilitate and intensify orgasms Puppo (2011). Also, touch and pain perception thresholds increase with the physiological arousal levels under stimulation with vibrations but not for thermal stimulation Gruenwald et al. (2007). It is also understood

that experiencing fear induces activity of the pelvic floor muscle van der Velde et al. (2001); Both et al. (2012) and tense pelvic floor muscles before or at the beginning of the penetrative act may then lead to decreased blood flow and lubrication Van Lunsen and Ramakers (2002); Binik et al. (2006). Hence, penetrative activities with no or little arousal or initial activity of the pelvic floor muscles due, e.g., to fear may cause vulvar pain Brauer et al. (2006); Farmer and Meston (2007); ter Kuile et al. (2010). Several models of the behaviour of the pelvic floor muscles focussing on childbirth (but not considering GPPD) are summarised in Li et al. (2010).

Unfortunately the models presented above do neither include all or even the majority of the variables involved in experiencing genital pain, nor describe the dynamics of these variables from quantitative perspectives. One of the few models that describe the interplay of several key variables as a dynamic model is derived in Varagnolo et al. (2017), where the variables form two distinct loops, referred to as the Circle Of Fear (COF) and Circle Of Pleasure (COP). The COF captures the facts that: *i)* pelvic muscle activity before or at the beginning of penetration may lead to pain; *ii)* fear induces muscular tension; and *iii)* inducing positive erotic stimuli may reduce fear. The COP is instead based on the Basson model of the female sexual response Basson (2000), and captures the facts that: *i)* the physiological arousal increases under sexual stimulation and when feeling subjectively aroused; *ii)* the subjective arousal increases with sexually stimulation and pleasurable physical sensations; and *iii)* physiological arousal affects the subjective arousal indirectly via the physical pleasure. However, the model in Varagnolo et al. (2017) is solely based on known cause-effect relationships in the medical literature and informed guesses from experts in the field. Moreover its objectives are to find a suitable deterministic mathematical and quantitative model that strikes a balance between being able to accurately model some known relationships and being simple enough to be mathematically analysable. However, the model in Varagnolo et al. (2017) is neither directly based on specific medical tests nor measurement data, and is hence not validated.

Consider then that the medical literature clearly suggests the existence of interplays between psychological and physiological responses in patients subject to vaginal dilation stimuli. For designing these stimuli in a quantitative way, there is thus the need for quantitative models of both these types of responses. As for the psychological response to vaginal dilation, a data driven model was created by our team (and proposed in Varagnolo et al. (2018)) using statistical analysis tools. Varagnolo et al. (2018) verified that clustering the psychological responses of patients in groups may lead, from a data-driven point of view, to models with a significantly improved statistical performance. Overall the paper thus suggested that casting the modelling of subjective pain/pleasure assessments in response to vaginal dilation stimuli as a support vector classification problem can lead to models having classification error performances in test sets up to 25%.

As for the data-driven modelling of the physiological responses part, a first step towards closing the knowledge gap was proposed in Knorn et al. (2018). Here, our team derived data-driven dynamical models of female response to vaginal dilation using time-series of pelvic floor pressure collected from healthy patients during ad-hoc medical trials. The work in Knorn et al. (2018) investigates which type of model and model order are suitable to accurately describe the recorded data both for individual patients' models and "average" models (i.e., models built considering data from several individuals).

1.2 Statement of contributions

This manuscript extends and completes Knorn et al. (2018) in the following ways: even though the data used in Knorn et al. (2018) was gathered from medical trials, where women watched different video clips (including low or high arousal sexual or non-sexual movies), the information of which video was watched when during the sessions was not used. This means that the models proposed in Knorn et al. (2018) ignore the psychological effects of watching different types of videos on the physiological response, in a sense thus neglecting potential statistical dependencies that may be present in the collected evidence.

The work presented here extends the results from Knorn et al. (2018) by also considering how the models derived for specific movie types differ. Further, we investigate the variations between models for certain movies compared to the variation between different patient models. In a sense, we focus on answering in a quantitative fashion the following questions: What are suitable model structures for this type of system? Are there differences in the physiological responses to sexual vs. to nonsexual movies? How uniformly do different people physiologically respond to the same movie? Given our vision of providing tools for designing personalised vaginal dilation patterns, we seek to answer the questions above by focusing specifically on models with control-oriented structures such as Hammerstein-Wiener and Nonlinear autoregressive exogenous (NARX) models, which have been shown to be suitable in other biomedical applications, see Bro and Medvedev (2017); Langdon et al. (2016).

1.3 Organization of the manuscript

The paper is organised as follows: Section 2 describes the experimental setup used to record the measurement data. Section 3 overviews the standard strategies of modelling generic muscular activity. The notation is clarified in Section 4. Sections 5 and 6 present our identification results. Section 7 closes the paper by drawing some qualitative and quantitative conclusions.

2. MEDICAL TRIALS SETUP

To derive quantitative dynamical models of how the pelvic floor muscles respond to forced vaginal dilation we use the dataset recorded at Maastricht University Hospital (described in more detail in Melles et al. (2018)). The data includes participants' responses to the gradual vaginal dilation forced by an inflatable balloon to be inserted at the introitus as described in Figure 1, called Vaginal Pressure Inducer (VPI). During the experiments, women also watched sequences of 5-minutes long erotic or non-erotic movies in the (tentatively) neutral environment shown in Figure 2.
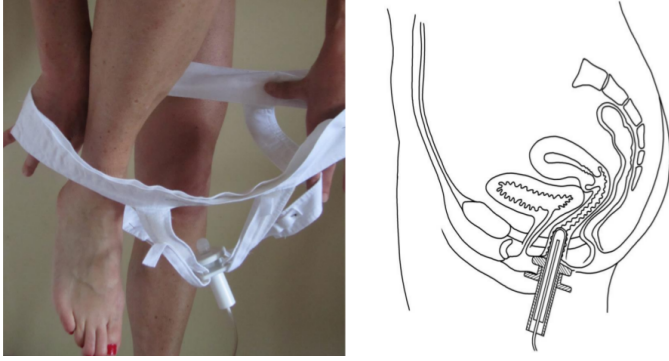
Fig. 1. Picture of the VPI (left) and schematic description of its usage (right). The balloon can be gradually expanded by filling it with water at body temperature by a pump; the length of the inflated area is up to 6 cm. When the balloon is filled, an outward omnidirectional pressure is given to the surrounding tissues.

Fig. 2. Photos of the room hosting the medical trials.

The study included data from 36 women without sexual problems, aged between 18 and 45 years, in a steady heterosexual relationship for at least 3 months, and being sexually active including coitus. Each woman participated in a single session where, while using the VPI and watching movie sequences, they recorded their perceived level of comfort/discomfort (on a scale from 0, maximum admittable discomfort, to 100, complete comfort) with an opportune slider. As soon as the pressure felt unbearable, participants could end the experiment and start the deflation of the balloon. The sessions started with the presentation of a neutral acclimatisation movie with pressure induction using the VPI. This was followed by showing one high-arousal sexual movie without inducing vaginal pressure, then followed by four randomised movies with inducing pressure. Every patient watched four movie clips in randomized order: Movie 1: an explicit erotic clip, meant to be sexual and inducing arousal. Movie 2: a clip of a heterosexual couple kissing, meant to be sexual but not inducing arousal. Movie 3: an excerpt from Roberto Benigni's "Life is beautiful", meant to be non-sexual and inducing arousal. [1] Movie 4: an excerpt from a documen-

---

[1] Note that the movie is intensely sad and is enhance expected to induce strong emotional reactions without being sexual.
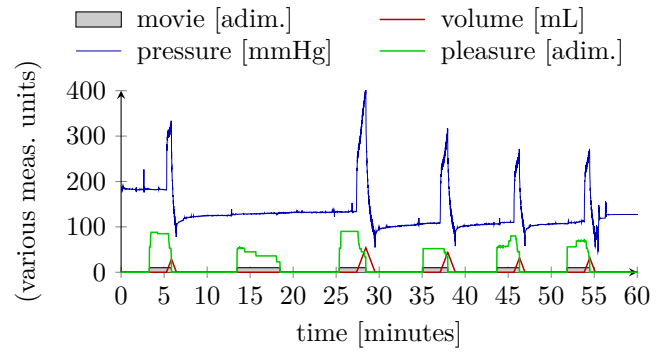

Fig. 3. Dataset from one patient. The six movie clips described above are in this case started at minutes 3, 13, 26, 35, 43 and 52. The VPI was inserted in the duct during the whole trial but inflated only while watching the movies (but the second one).

tary on the life of zebras, meant to be non-sexual and not inducing arousal.

During the experiment, the induced pressure in the water was measured at the pump as an indirect measure of the pelvic floor muscle activity. Further, the reported pleasure levels were recorded as well as when patients stopped the experiment to enforce the deflation of the balloon. A typical data set is shown in Figure 3.

3. MODELLING THE PHYSIOLOGICAL RESPONSE TO VAGINAL DILATION: CHOICE OF THE SUITABLE MODEL STRUCTURES

We are interested in using the measurement data recorded in the test described in Section 2 to model the psycho-physiological system having, as inputs:

- psychological assessments in the form of recorded pleasure levels,
- psychological stimuli in the form of different movies types,
- physiological stimuli in the form of enforced vaginal (volumetric) dilation levels,

and as outputs the aggregated muscular pressure exerted by the pelvic floor muscles.

This modelling problem is similar to the general case of relating muscular stimulation levels with the corresponding pressure (or force) outputs, for which many different generic models of different complexities exist. These include *physiologically based models*, that relate stimuli and corresponding forces as interactions of the fibers at a microscopic level, Huxley (1957), *Hill-type models*, that relate stimulation levels and corresponding forces through mechanically-inspired concepts, Hill (1977), and *black-box models*, that derive input-output relations from numerical evidence. Note that physiologically based models or Hill-type models require the additional data such as (1) measurements of the muscular stimulation signals or other signals that are known to be correlated (e.g., Electromyography (EMG) levels); (2) measurements of the force or pressure exerted by a specific set of muscular fibers; or (3) measurements of the mechanical parameters of the muscular fibers (such as thickness and length). However,
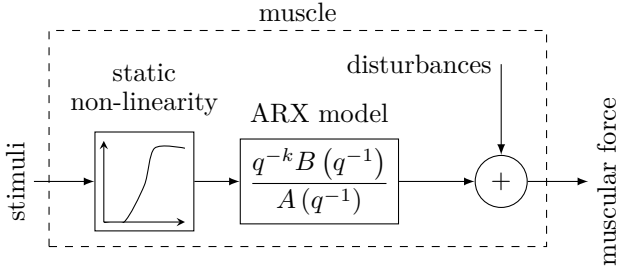
Fig. 4. Graphical representation of an Hammerstein model of muscular dynamics Hunt et al. (1998). Hammerstein-Wiener models generalise further these types of models by adding a further static nonlinear map after the ARX transfer function.

since such information is not included in the data used for this work, we thus follow a purely data-driven approach.

In the literature on black-box models for muscular dynamics, the most common strategies use *Hammerstein-Wiener* or *NARX* models, including Neural network (NN) and fuzzy models. Since physiological models of muscular dynamics are typically non-linear, non-linear identification approaches tend to provide better results than linear ones.

Hammerstein models, see Figure 4, include a static nonlinear map (called the *static recruitment*), an Autoregressive exogenous (ARX) model, and an additive disturbance that may account for temporary effects like fatigue and that can be modelled through another additional transfer function.

NARX models (Ljung, 1997, Chap. 5) can be described in the general form

$$y(t) = f\Big(y(t-1), \ldots, y(t-T), u(t-1), \ldots, u(t-T), \theta\Big) + d(t)$$
(1)

where y is the output, u the input, and d disturbances of the system. There exists a vast literature on how to determine both the structure of $f(\cdot)$ and the best set of parameters θ. It is also known that the problem of selecting the structure of f is a difficult task specially when the size of the available dataset is small. For an example of modelling muscular dynamics through NARX approaches see, e.g., Previdi (2002).

We notice that machine-learning inspired approaches such as NN and fuzzy models may achieve great generalization capabilities when modeling input-output muscle dynamics. Their main drawback is though that they are more difficult to be used for automatic control purposes. On the other hand, Hammerstein-Wiener and more control-oriented NARX models have also been proven to be capable of high approximation capabilities for various medical applications (e.g., Bro and Medvedev (2017); Langdon et al. (2016)). In the paper we thus explore the approximation capabilities of Hammerstein-Wiener models with different nonlinearities and some NARX strategies for the available dataset.

A special discussion shall then be posed to describe how to account for the psychological components that we aim at using as inputs, i.e., the recorded pleasure levels and the different movies types that patients watch during the medical trial.

As described in our literature review, it is well known that the human psycho-physiological sexual response is heavily dependent on both physiological measures and external stimuli. While some external factors such as stimulating music, books, movies, conversations etc may lead to an increase in physiological and subjective sexual arousal, others such as stress, hostile partner responses or relational conflicts may reduce arousal or act as turn-off factors.

In the experiments conducted at Maastricht university hospital and considered in this paper, women watched different types of short movie sequences. Since they include sexual movies (high and low arousal) and nonsexual movies (high and low arousal), it is reasonable to assume that they also influence the pelvic floor muscle response to vaginal stimuli. Even though the specific reaction of individual women to the same movies might differ, we expect that specific types of movies tend to induce similar types of emotional responses.

In this paper we thus aim also to investigate whether different types of movies induce different reactions by studying whether models trained for a specific movie type have higher prediction capabilities than models trained for an individual patient. In other words, we investigate whether the type of movie that has been seen by the patients should be considered an effective external stimulus or not.

We finally note this type of external influence can be included in the model in different ways, i.e., as a (partly unknown) external disturbance, as a (control) input, or simply as a factor that influences the parameters of the system. In any case, assessing the amplitude of the effect induced by movies is important as it is an evaluation of how much we expect physiological responses to depend on external factors for this specific type of framework.

We will hence investigate the prediction capabilities for individual models in Section 5, i.e., one model per patient, and the prediction capabilities per movie type in Section 6, i.e., one model per movie type.

4. NOTATION

For notational clarity, let $i = 1, \ldots, 36$ denote the patient ID, and $\mathcal{D}_i$ its associated dataset. Every $\mathcal{D}_i$, as described in Section 2, is composed by 4 time-series signals. Letting t denote the time index, and incidentally noting that all the various signals have sampling periods equal to 1 second, in more details the collected signals are, for each patient i:

- $m_i(t)$, a non-negative integer in $\{0, 1, 2, 3, 4\}$ indicating if a patient i was watching a movie at time t ($m_i(t) \neq 0$) or not ($m_i(t) = 0$). If $m_i(t) \neq 0$ then the signal indicates also which movie type was being shown to the patient;
- $\ell_i(t)$, a non-negative integer in $\{0, \ldots, 100\}$, indicating the perceived pleasure level of patient i at time t. Note that this signal is nonzero only when a movie is being played;
- $v_i(t)$, indicating the volume of the VPI at time t for patient i;
- $p_i(t)$, indicating the measured aggregated pelvic floor muscles pressure at time t for patient i.

We are then interested in learning models of the generic form

$$p_i\,(t+1) = \phi_i\,(p_i(t),\ldots,p_i(t-T),\ldots$$
$$v_i(t),\ldots,v_i(t-T),\ell_i(t),\ldots,\ell_i(t-T),\ldots \quad (2)$$
$$m_i(t),\ldots,m_i(t-T)\;;\;\theta_i)\,.$$

Note that the subscript i in (2) implicitly indicates a notation referring to an individual model, i.e., a function ϕ_i that captures how the specific patient i responds to stimuli and that has been learned by using data relative to patient i. When we instead refer to a model that has been trained using data from several patients we remove the subscript i and indicate the model just with ϕ. In a similar fashion, rather than considering individual models for individual patients, we can also consider individual models for individual *movies types*, i.e., join the responses of several patients to the same movie $m \in \{1,\ldots,4\}$ and use this gathered data in order to train a model that refers to that specific movie m. In this case we use the notation ϕ_m, with m reminiscent of "movie" instead of i, reminiscent of "individual" (patient).

The quantities especially of interest for our purposes are:

- the functional structures of ϕ_i, ϕ or ϕ_m, assumed to be selectable within a finite set of plausible functional structures denoted with

$$\Phi := \left\{ \phi^{(1)},\ldots,\phi^{(M)} \right\}; \quad (3)$$

- the model order T_i, assumed to be selectable within a finite set of plausible orders denoted with

$$\mathcal{T} := \left\{ T^{(1)},\ldots,T^{(N)} \right\}; \quad (4)$$

- the vector of model parameters θ_i whose dimension depends on which structure ϕ_i and order T_i is used.

As for the set of plausible functional structures Φ we consider the set of available alternative choices when using Matlab's 2018b `system identification toolbox 9.9` – in practice, Hammerstein-Wiener models with different structures for the input and output nonlinearities, plus wavelet, tree-partitioning, and sigmoid NARX models. Notice that neural networks were not taken into consideration, since our focus here is on structures whose representations can give intuitive knowledge to physicians and in general people not coming from engineering fields.

5. RESULTS – INDIVIDUAL PATIENTS MODELS

In this section, we present results relating to predicting the responses of patients using individual models. Recall that during the collection of the generic dataset $\mathcal{D}_i$, the related patient i was exposed to a total of 6 movies (see, e.g., Figure 3) in total, of which the last 4 movie clips are the four movie types (high arousal sexual, low arousal sexual, high arousal nonsexual and low arousal, nonsexual) in randomised order. Since this implies that for each patient only *one* set of data, watching each movie type only *once*, is available, the movie type is ignored when considering individual models. In case individual models, that also take the movie type into account, should be derived, several of these data sets per patient would be needed.

For practical reasons, in the following we neglect the data for the first two movies played. Indeed in the initial part of the trial, each patient acclimatizes with the experiment during the first movie, and the VPI balloon is not inflated during the second movie. An "individualized" learning process for every specific patient may be performed implementing the following pseudo-code:

(1) divide the four remaining movies played within the dataset $\mathcal{D}_i$ into two separate datasets, namely create a training set $\mathcal{D}_i^{\text{train}}$ comprising the responses to movies 3 and 4 (referring to the order in which they were played but neglecting the movie type) and a test set $\mathcal{D}_i^{\text{test}}$ comprising the remainder of the available evidence for patient i;

(2) for every potential structure $\phi_i^{(j)} \in \Phi$ and admissible model order $\mathcal{T}$, learn the model using patient i's training set $\mathcal{D}_i^{\text{train}}$;

(3) select the best model structure, order and parameters for patient i as that triplet that leads to the best fit in the test set $\mathcal{D}_i^{\text{test}}$.

In general, this strategy leads to individual models that differ in their structures j and/or orders $\mathcal{T}$. Different model structures, nonetheless, make the task of comparing and grouping different patients difficult. Using only one model structure and model order for learning the individual model parameters, on the other hand, has the potential drawback of reducing the generalization capabilities of the estimated models.

To evaluate this trade-off quantitatively there is the need for solving the ancillary question of how to select the model structure and order among the alternative competing choices. We thus consider the following strategy (note that the superscripts s and o are mnemonics model *structures* and model *orders*):

(1) for every patient i learn $M \cdot N$ models (i.e., one for each couple model structure / order $(\phi^{(s)}, T^{(o)})$ of potential alternative model structures and order choices) using the individual training set $\mathcal{D}_i^{\text{train}}$. This means learning for each patient i $M \cdot N$ different parameters vectors $\widehat{\theta}_i^{(s,o)}$;

(2) for every learned $\widehat{\theta}_i^{(s,o)}$ (thus for every patient i, model structure s and model order o) compute the simulated pressure

$$\widehat{\boldsymbol{p}}_i^{(s,o)} := \phi_i^{(s,o)}\left(\widehat{\boldsymbol{p}}_i^{(s,o)}, \boldsymbol{v}_i, \boldsymbol{\ell}_i \;;\; \widehat{\theta}_i^{(s,o)}\right) \quad (5)$$

where we tacitly let the signals $\boldsymbol{v}_i$ and $\boldsymbol{\ell}_i$ in (5) belong to the test set $\mathcal{D}_i^{\text{test}}$, we omit writing time delays, and we assume that the initial conditions for initializing the simulations are known and set to be equal to the initial measurements in the test set;

(3) for every simulated pressure $\widehat{\boldsymbol{p}}_i^{(s,o)}$ we compute its fit index in the test set as

$$\mathcal{F}_i^{(s,o)} := 100 \cdot \left(1 - \frac{\left\| \boldsymbol{p}_i - \widehat{\boldsymbol{p}}_i^{(s,o)} \right\|}{\left\| \boldsymbol{p}_i - \text{mean}(\boldsymbol{p}_i) \right\|} \right); \quad (6)$$

(4) for every couple (s,o) of potential model structure and model order we compute its average fit over the set of all the patients, i.e., compute

$$\overline{\mathcal{F}}^{(s,o)} := \frac{1}{36} \sum_{i=1}^{36} \mathcal{F}_i^{(s,o)}. \quad (7)$$
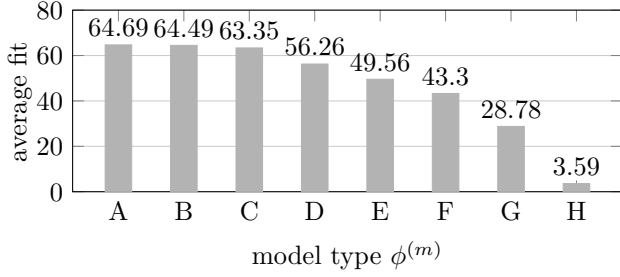
Fig. 5. Average fits on the test sets for various potential model structures (the model order being here implicitly assumed to be, for each model structure $\phi^{(s)}$, that one that maximizes $\overline{\mathcal{F}}^{(s,o)}$ over the potential orders o). Legend: A = Hammerstein-Wiener (HW) piecewise (pw).linear - pw.linear, B = HW pw.linear-saturation, C = HW poly.-poly., D = sigmoidnet, E = ARMAX, F = HW pw.linear-deadzone, G = wavenet, H = treepartition. To complement the information, the standard deviations of the fit indexes associated to the first three model structures are, respectively, 23.89, 24.02, and 25.68.

In this way we check the generalization performance of that specific model structure and order;

(5) we finally select as best model structure s^* and best model order o^* that couple s, o that is associated to the highest average fit $\overline{\mathcal{F}}^{(s,o)}$.

Notice that since this procedure is reminiscent of a cross-validation approach we do not employ information criteria like Akaike or Bayesian ones for penalizing higher model orders.

Figure 5 shows then a summary of the results obtained following the procedure above. More precisely, the figure reports the average fits $\overline{\mathcal{F}}^{(s,o)}$ for a subset of potential model structures; for each structure we then plot for simplicity only the data associated to its best model order. Interestingly, the 3 model structures returning the best average fits are all Hammerstein-Wiener models and all with similar functional structures in the input and output nonlinearities. This seems to indicate that for our specific framework of modelling pelvic floor muscular pressure as a function of vaginal dilation we recover the same functional structures that have been proposed in the literature for modelling generic muscular force as a function of EMG levels, see for instance Hunt et al. (1998).

In the following we thus assume that the "best model structure" s^* to work with is a Hammerstein Wiener model with piecewise linear input and output nonlinearities, as indicated by the results on both average and standard deviation of the fit performance in Figure 5. In other words, in the remainder of the section we fix the structure and order of the models for all patients to be this specific one, but at the same time let the parameters of the individual models be potentially different and depending on the specific patient i. We then quantify how much the performance of the individual estimators $\widehat{\theta}_i^{(s^*,o^*)}$ may vary depending on the particular patient under consideration, and thus analyze the spread of the statistical performance of $\mathcal{F}_i^{(s^*,o^*)}$ over i. Instrumental to this evaluation, we
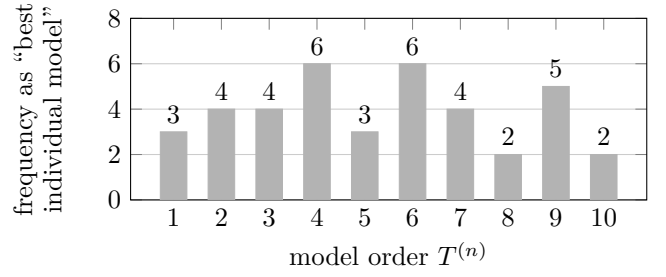

Fig. 6. Histogram of how many patients have a specific model order as their best individual one, as determined through strategy S1 above.

first check the sensitivity of the fit indexes on the model order while assuming the structure fixed. Consider thus the alternative strategy:

S1) select as the "best individual model order" that individual order n_i^* that maximizes the individual fit $\mathcal{F}_i^{(s^*,o)}$ over o;

Figure 6 refers to strategy S1 above and plots how many patients had a certain individual model order as their best one. Unfortunately the plot does not give clear indications on what may be a suitable o^* for strategy S1, in the sense that the curve is neither unimodal nor with a small overall spread. Considering however not only how many patients have a specific model order as best fit (see Figure 6) but also the actual fit values, the best order seems to be $o^* = 4$.

Figure 7, instead, compares the histograms of two sets of fit indexes: the set of indexes $\mathcal{F}_i^{(s^*,o^*)}$ obtained by fixing the model structure and order to be the same for the various patients (named "fixed order" in the figure), and the set of indexes $\mathcal{F}_i^{(s^*,n_i^*)}$ obtained using strategy S1, i.e., where the order of the models are individual variables (named "individual order" in the figure). As expected, the best fits can be obtained when allowing individual choices of model orders, since this guarantees more flexibility. The trade-off becomes thus the following: from practical reasons, allowing individual model orders might help getting models with better prediction capabilities. Restricting the models to have the same orders on the other hand allows to compare different estimated parameters $\widehat{\theta}_i$ for different patients, and this in its turn enables introducing algorithms for grouping and clustering the patients based on opportune geometric distances among their learned parameters.

For the sake of completeness, we finally consider the performance that can be obtained fixing the very same model structure s and order o for all the various patients (and, more precisely, using that structure $s^* =$ HW pw.lin and order $o^* = 4$ determined when computing Figure 5). We thus report in Figure 8 the simulation results for the patient associated to the worst fit against the simulations relative to the patient with the best fit. The simulation results shows that for the best case the learned model is qualitatively able to reproduce the features of the measured time series. For the worst case, instead, the simulated pressure has very poor approximation capabilities. This indicates that this general model is not suitable for some
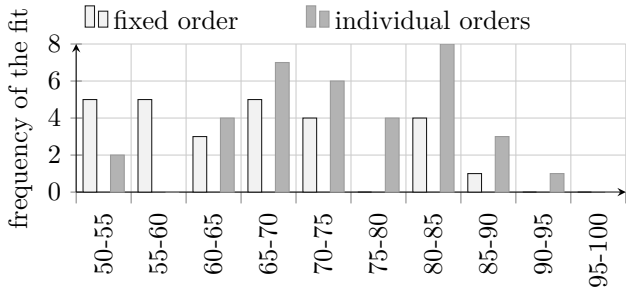
potential fit values of the $\mathcal{F}_i^{(s^*,o^*)}$'s and the $\mathcal{F}_i^{(s^*,n_i^*)}$'s

Fig. 7. Histograms showing for how many patients a certain fit value has been reached in the test set. We first consider the indexes $\mathcal{F}_i^{(s^*,o^*)}$ corresponding to using the model structure and order with the best average fit indicated in Figure 5 to simulate the various individual patients (white, corresponding to a Hammerstein-Wiener (HW) pw.linear-pw.linear with order 4), and then consider the indexes $\mathcal{F}_i^{(s^*,n_i^*)}$, obtained by using a fixed model structure but allowing individual model orders vary for each patient (gray).
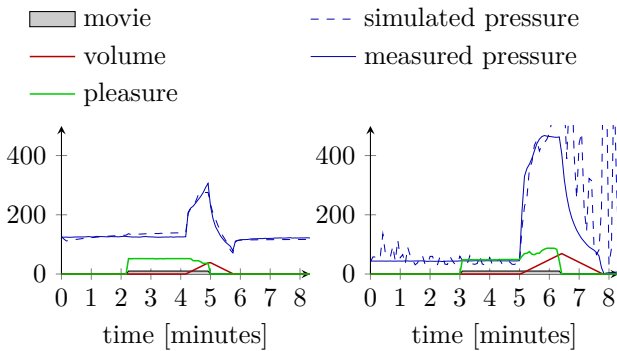

Fig. 8. Comparison of the simulation results for the patients associated to the best fit (left) and worst fit (right), choosing the model structure and order s^* and o^* with the best fit according to Figure 5. A summary of the modelling performance for the other 41 patients is implicitly included in Figure 7.

patients. It indicates that for modelling the responses of these persons there is the likely need of considering additional data, factors or disturbances.

6. RESULTS – MOVIE-DEPENDENT MODELS

The previous Section 5 discussed how different model types (in the sense of structures and orders) are able to fit the data, and tried to answer the question "what are suitable model structures for this type of system?".

The learning approaches used in Section 5, however, ignored the information about which type of movie the patients were watching (that are, for the sake of completeness, of 4 different types, combining the dichotomies high-arousal / low arousal, sexual / non-sexual). The need for ignoring this information is due to the lack of enough measurements: each patient watched exactly every type of such movies just once, so that it is not possible to

perform the classical "learn on training data, assess on test data" since it is not possible to divide the available data in meaningful train and test datasets.

Disregarding the fact that watching different movie types may potentially induce psychological effects is, however, a clear drawback of the models proposed in Section 5, specially considering that the medical literature indicates that psychological and physiological responses are tied.

To fill this gap in this section, we focus thus on answering two specific questions: are there differences in the physiological responses to different movie types? And how uniformly do different people physiologically respond to the same movie type?

To do so we first divide each dataset (like the one in Figure 3) in 6 segments so that each segment refers to one single and specific movie. To do so, we arbitrarily defined these segments in a way that they temporally start at:

- time 0 for the first segment;
- two minutes after the VPI has completely deflated for the second, fourth, fifth and sixth segment;
- two minutes after the second movie has finished for the third segment.

We moreover let these segments finish either one minute before the beginning of the new segment or at the end of the experiment. Qualitatively, thus, we include in each movie segment the various signals before the movie starts up to a few minutes after that movie ends.

Given the above segmentation criterion, we then create four different training sets and four test datasets $\mathcal{D}_m^{\text{train}}$ and $\mathcal{D}_m^{\text{test}}$, with m reminiscent of "movie", starting from the "individual patients i" training and test datasets $\mathcal{D}_i^{\text{train}}$ and $\mathcal{D}_i^{\text{test}}$ considered in Section 5. More precisely, $\mathcal{D}_m^{\text{train}}$ contains the segments of various $\mathcal{D}_i^{\text{train}}$, that correspond to the specific movie type m, for the first half of the patients, i.e., for i between 1 and 18. Then, the responses of the second half of the patients, i.e., for i between 19 to 36, when watching movie type m form $\mathcal{D}_m^{\text{test}}$.

Importantly, the operation of merging data segments relative to one specific movie but also relative to different patients hides a technical difficulty: patients' baseline pressures are both individual and time-dependent. I.e., inspecting Figure 3 (that is representative of what happens also with all the other various patients) and comparing it to the data from other patients we can notice two distinct phenomena:

(1) the vaginal pressure during resting periods (i.e., while the VPI is completely deflated) can be considered a plateau (i.e., relatively stable and flat in time). Consecutive resting periods tend to be characterized by lower and lower plateau levels;
(2) different people experience very different resting vaginal pressure plateau levels.

Relative to the first phenomenon, we note that indeed the first plateau is sensibly higher than the consecutive ones in virtually all the analysed patients. An explanation supported also by interviews with patients is that at the beginning they were subject to different levels of

anticipatory anxiety, that then vanish as the patients get more and more relaxed as the experiment goes on.

The two phenomena described above imply that joining data from different individuals requires an ad-hoc detrending strategy. Our strategy has been to implement, for each of these identified segments, the following pseudo-code:

(1) first detrend the pressure signal within the segment (call it $p_i^s(t)$) by removing the value of its first sample. In this way the new signal $\overline{p}_i^s(t) := p_i^s(t) - p_i^s(0)$ will start from zero;

(2) then consider the value of the last sample for this detrended pressure signal, say $\overline{p}_i^s(T)$. Our aim is to modify $\overline{p}_i^s(t)$ so to arrive at a new signal $\overline{\overline{p}}_i^s(t)$ that is s.t. both $\overline{\overline{p}}_i^s(0)$ and $\overline{\overline{p}}_i^s(T)$ are equal to 0;

(3) to do so, we note that - inspecting the datasets - all the segments present vaginal pressure traces that share the same qualitative behavior: patients physiologically respond to an inflating VPI by first exerting an increasing vaginal pressure, but as the VPI completely deflates the patients experience a brief over-relaxation that lasts few seconds. After this "overshoot" the patient then increases her vaginal pressure, and a new plateau begins. To compute a $\overline{\overline{p}}_i^s(t)$ that is s.t. both $\overline{\overline{p}}_i^s(0)$ and $\overline{\overline{p}}_i^s(T)$ are equal to 0 but that at the same time retains statistical information of how patients dynamically respond to an increasing volumetric vaginal stimulus we then find this "overshoot" and adjust the samples after this overshoot so that eventually the last sample of $\overline{\overline{p}}_i^s(t)$ will be zero.

In a sense, thus, the final $\overline{\overline{p}}_i^s(t)$'s considered in the learning phases are structurally equal to the original $p_i^s(t)$'s up to the negative overshoot defined above, and then after that differ from the original signal by a typically small bias. Among the potential choices for performing this detrending operation we choose the here defined one because it does not alter the statistical information of how patients dynamically respond to an increasing volumetric vaginal stimulus. It alters though the information relative to how patients respond to the transition inflated → completely deflated. Our detrending strategy though modifies only the gain associated to this operation, while preserving the time constants. The effects of different detrending choices have not yet been evaluated and are left to future work.

Assuming that the above detrending operation has been implemented in all movie segments and that the datasets $\mathcal{D}_m^{\text{train}}$ and $\mathcal{D}_m^{\text{test}}$ have been computed, we then can repeat the same strategy implemented in Section 5, i.e.,

- for every potential structure $\phi_m^{(j)} \in \Phi$ and admissible model order $\mathcal{T}$ we can learn the movie-type m's model using the training set $\mathcal{D}_m^{\text{train}}$;
- we can then select the best model structure, order and parameters for the movie-type m as that triplet that leads to the best fit in the test set $\mathcal{D}_m^{\text{test}}$.

This means that for each m we can select the best model structure s^* and best model order o^*, i.e., the couple s, o that is associated to the highest average fit $\overline{\mathcal{F}}_m^{(s,o)}$, using the same notation as above. Performing this check is instrumental to verify that the models obtained for indi-
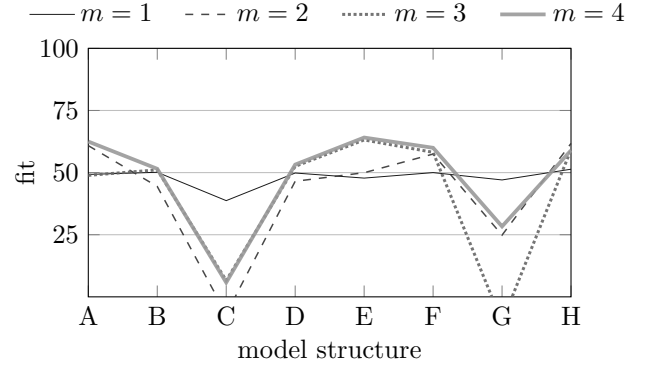

Fig. 9. Fits on the test sets for different potential model structures (the model order being here implicitly assumed to be, for each model structure $\phi^{(s)}$, that one that maximizes $\overline{\mathcal{F}}_m^{(s,o)}$ over the potential orders o) for the four different movies types. Legend: A = HW pw.linear - pw.linear, B = HW pw.linear-saturation, C = HW poly.-poly., D = sigmoidnet, E = ARMAX, F = HW pw.linear-deadzone, G = wavenet, H = treepartition. The movie indexes refer to 1 = "high arousal, sexual", 2 = "low arousal, sexual", 3 = "high arousal, nonsexual" and 4 = "low/no arousal, nonsexual" as described in Section 2.

vidual patients and for individual movies are structurally the same.

Figure 9 shows that once again HW models with piecewise linear nonlinearities lead to the best approximation capabilities, with fit indexes in test sets in the order of 60%. This in a sense is expected and also implicitly reassures that the detrending strategy introduced above does not introduce radical changes from a signals-information content point of view.

We then inspect the variability of the (test-set) fit performance in simulating the responses associated to specific movies of the various individual patients in the following sense: Assume that the best model (in the sense defined above) trained for movie m on the dataset $\mathcal{D}_m^{\text{train}}$ is ϕ_m^*. If then the associated test set $\mathcal{D}_m^{\text{test}}$ is composed by say K traces from K different individuals, then ϕ_m^* will have in general K different fit performances in simulating these K different traces. The related evidence collected in our medical trials is then shown in Figure 10.

Figure 10 shows that the greatest variability in terms of performance indexes happens for movie type 1, i.e., explicit erotic clips. This graph is in a sense a quantitative assessment of how much different persons have different responses to this type of movies. The smallest variability is happens instead for movie type 3, i.e., Benigni's "Life is beautiful". The evidence thus seems to suggest that (even if wary of the statistical significance of the numerical results above) from structural points of view, the measured physiological responses to both sexual and nonsexual movies are captured through HW models of the same type. However, movies with sexual content (specially 1) may lead to more variegate individual physiological responses than non-sexual movies (3 and 4). We guess that this higher variability may be due to more different emotional responses to these movies across the spectrum
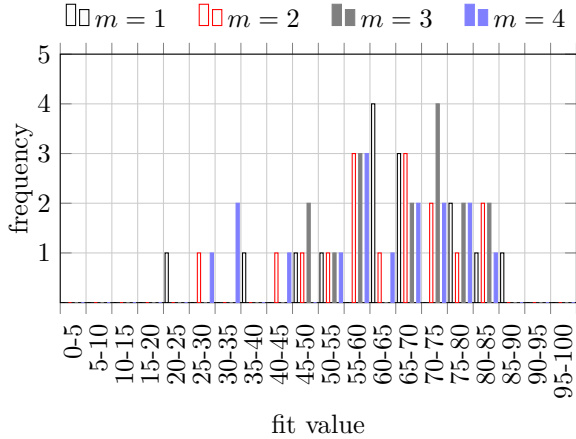
Fig. 10. Variability of the fit performance in the test set of the optimal individual models ϕ_m^*, $m = 1, \ldots, 4$, in simulating the responses associated to the specific movies of the various individual patients. The movie types indicate 1 = "high arousal, sexual", 2 = "low arousal, sexual", 3 = "high arousal, nonsexual" and 4 = "low/no arousal, nonsexual", see also Section 2.

of the patients, and that these differences eventually cause also higher variations in the physiological responses. However, we do not have data supporting this claim: checking this would indeed require opportune statistical analyses of subjective questionnaires and qualitative self-assessments of the users experiences, well beyond the scope of this paper.

We finally quantitatively assess how different the four individual-movie models $\phi_{m=1}^*, \ldots, \phi_{m=4}^*$ are by comparing their performance in simulating the four test sets $\mathcal{D}_{m=1}^{\text{test}}, \ldots, \mathcal{D}_{m=4}^{\text{test}}$ (and, as a cross-check, the training sets $\mathcal{D}_{m=1}^{\text{train}}, \ldots, \mathcal{D}_{m=4}^{\text{train}}$). The results are summarized in Tables 1 and 2, whose generic (m', m'') element represents how well the model learned for movie m' is able to simulate the test or training set, respectively, associated to movie m''.

	$\mathcal{D}_{m=1}^{\text{test}}$	$\mathcal{D}_{m=2}^{\text{test}}$	$\mathcal{D}_{m=3}^{\text{test}}$	$\mathcal{D}_{m=4}^{\text{test}}$
$\phi_{m=1}^*$	51.36	63.50	59.13	63.23
$\phi_{m=2}^*$	50.93	61.74	60.55	66.52
$\phi_{m=3}^*$	46.84	65.92	63.14	65.48
$\phi_{m=4}^*$	43.70	57.77	53.57	64.11

Table 1. Fit performance of the individual-movie models ϕ_m^* in simulating the test sets $\mathcal{D}_m^{\text{test}}$.

	$\mathcal{D}_{m=1}^{\text{train}}$	$\mathcal{D}_{m=2}^{\text{train}}$	$\mathcal{D}_{m=3}^{\text{train}}$	$\mathcal{D}_{m=4}^{\text{train}}$
$\phi_{m=1}^*$	58.82	48.29	58.64	57.89
$\phi_{m=2}^*$	55.93	44.71	55.84	54.72
$\phi_{m=3}^*$	59.22	50.63	72.68	68.17
$\phi_{m=4}^*$	61.03	43.16	58.25	60.64

Table 2. Fit performance of the individual-movie models ϕ_m^* in simulating the training sets $\mathcal{D}_m^{\text{train}}$.

Tables 1 and 2 show the training and test fits for the different models per movie type and data sets. As indicated above, the training sets per movie type are formed by combining the corresponding responses of the first half of the patients whereas the responses of the second half of the patients was used to form the corresponding test set.

It can be seen that the training and test fits are comparable; this, together with leave-one-out cross validations (Hastie et al., 2008, Chap. 7.4) that are omitted for the sake of brevity, indicate that no overfitting is likely to have happened.

We then notice that the fit values between the various training and test sets vary more when the movie segments have sexual content rather than when the segments are without sexual content. A possible explanation may be that (as elaborated also above) sexual content may lead to more varied physiological responses, and this increases the variability of the results (even if we cannot however support this intuition with the data that we have collected in our trials). This is specially noticeable when realizing that the performance of the various learned models ϕ_m^* in simulating $\mathcal{D}_{m=1}^{\text{train}}$ (i.e., the responses associated to the erotic sexual clips) tend to be lower than in simulating the other $\mathcal{D}_m^{\text{train}}$'s. This in a sense confirms and reinforces the intuition developed before that movies with sexual content lead to individual physiological responses that are more difficult to be captured.

This seems to indicate that it is important to know if a patient is watching a sexual movie when identifying her physiological responses, while knowing if it is high arousal or low arousal seems to bring only minor information. For this reason we believe that the subjective pleasure indicator eventually cannot be considered as a signal that implicitly incorporates the information of the movie type, and shall be always used for identification purposes.

7. CONCLUSIONS

This paper presents results on data-driven modelling of the pelvic floor muscle dynamics for healthy patients undergoing vaginal dilation exercises through an inflatable balloon while being exposed to different movie types. For the available datasets, we identified that the pressure dynamics are best modelled as a Hammerstein-Wiener model with piecewise linear input output maps, a fact that is reminiscent of similar results in the medical literature dedicated to numerically modelling the dynamics of muscular pressure as a function of EMG levels.

In the paper we specifically focus on understanding what are the effects from a system identification point of view of enforcing the model type and model order for capturing the dynamics of all the various patients. We indeed aimed at checking whether different patients share dynamics with similar functional structures, and found results that are partially contradictory: Even if, as said above, Hammerstein-Wiener structures seemed to capture the collected evidence for all various patients, we have not been able to find a common order for the linear blocks of the various patients that led to satisfactory approximation capabilities for every patient. This is not ideal from a modelling perspective, since having the same model structure but different model orders for different patients prevents being able to compare (and thus group) patients by means of comparing (and grouping) their estimated parameters.

Another important quantitative result that we found is about the effect of being exposed to different movies types while being subject to vaginal dilation stimuli. Testing the

generalization capabilities of our learned models it turned out that patients' physiological responses while watching movies with explicit erotic content are more variegated than when watching non-erotic movies. This is implicitly important for the overall aim of this research line, that intends answering the problem of how to personalize vaginal dilation patterns by building dynamical models that are control-oriented.

The intuition is that personalizing the dilation patterns requires predictive models, i.e., models that can accurately forecast, what will be the short- and long-term effects of applying specific dilation patterns to a specific patient in a specific condition. Our findings suggest then that finding predictive models of how patients' physiologically respond to movies with explicit erotic content is more difficult. Treatments are at the same time sometimes associated to these types of movies, since they may help to achieve physiological and psychological arousal, that are known in their turn to help some patients to experience better appraisal of the treatment themselves.

Our main conclusion is thus that models, which allow a better prediction of the pelvic floor muscle response or give interpretable insights into the underlying relation between the variables, would further ease the derivation of personalised dilation patterns. But for this it seems that there is the need for models that incorporate further research on the underlying physiological and psychological responses as well as dedicated medical tests, that capture more data of interest and are not focused, as did here, on finding connections among few of the variables that are involved in the system.

REFERENCES

Basson, R. (2000). The Female Sexual Response: A Different Model. *Journal of Sex & Marital Therapy*, 26(1).

Bergeron, S., Khalifé, S., Glazer, H. I., and Binik, Y. (2008). Surgical and Behavioral Treatments for Vestibulodynia: Two-and-One-Half-Year Follow-up and Predictors of Outcome. *Obstetrics & Gynecology*, 111(1):159–166.

Binik, Y. M., Bergeron, S., and Khalifé, S. (2006). Dyspareunia and vaginismus: so-called sexual pain. In *Principles and practice of sex therapy*, chapter 5, pages 124–156. The Guilford Press.

Both, S., van Lunsen, R., Weijenborg, P., and Laan, E. (2012). A New Device for Simultaneous Measurement of Pelvic Floor Muscle Activity and Vaginal Blood Flow: A Test in a Nonclinical Sample. *The Journal of Sexual Medicine*, 9(11):2888–2902.

Boyer, S. C. (2009). *The relationship between pain sensitivity and vasocongestion due to sexual arousal in women with provoked vestibulodynia*. PhD thesis, Queen's University.

Brauer, M., Laan, E., and ter Kuile, M. M. (2006). Sexual Arousal in Women with Superficial Dyspareunia. *Archives of Sexual Behavior*, 35(2):187–196.

Bro, V. and Medvedev, A. (2017). Nonlinear dynamics of the human smooth pursuit system in health and disease: Model structure and parameter estimation. In *Conference on Decision and Control (CDC)*.

Farmer, M. A. and Meston, C. M. (2007). Predictors of genital pain in young women. *Archives of sexual behavior*, 36(6):831–43.

Goldstein, A., Pukall, C., and Goldstein, I. (2009). *Female Sexual Pain Disorders*. John Wiley & Sons.

Goldstein, A. T., Klingman, D., Christopher, K., Johnson, C., and Marinoff, S. C. (2011). Surgical Treatment of Vulvar Vestibulitis Syndrome: Outcome Assessment Derived from a Postoperative Questionnaire. *The Journal of Sexual Medicine*, 3(5):923–931.

Gruenwald, I., Lowenstein, L., Gartman, I., and Vardi, Y. (2007). Physiological changes in female genital sensation during sexual stimulation. *Journal of Sexual Medicine*, 4(2):390–394.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning*. Springer.

Hill, T. L. (1977). *Free Energy Transduction in Biology*. Academic, New York.

Hunt, K., Munih, M., Donaldson, N., and Barr, F. (1998). Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle. *Biomedical Engineering, IEEE Transactions on*, 45(8):998–1009.

Huxley, A. F. (1957). Muscle structure and theories of contraction. *Progress in Biophysics and Biophysical Chemistry*, 7:255–318.

Knorn, S., Varagnolo, D., Oliver-Chiva, E., Melles, R., and Dewitte, M. (2018). Data-driven modelling of pelvic floor muscles dynamics. In *IFAC Symposium on Biological and Medical Systems (BMS)*.

Langdon, R., Docherty, P., Chiew, Y.-S., Möller, K., and Chase, J. (2016). Use of basis functions within a nonlinear autoregressive model of pulmonary mechanics. *Biomedical Signal Processing and Control*, 27.

Levin, R. J. (2002). The Physiology of Sexual Arousal in the Human Female: A Recreational and Procreational Synthesis. *Archives of Sexual Behavior*, 31(5):405–411.

Li, X., Kruger, J., Nash, M., and Nielson, P. (2010). Modeling childbirth: Elucidating the mechanisms of labor. *Systems Biology and Medicine*, 2(4).

Ljung, L. (1997). *System Identification: Theory for the User, 2nd Edition*. Prentice Hall.

Melles, R., Dewitte, M., ter Kuile, M. M., Bonnemayer, C., and Peters, M. (2018). The vaginal pressure inducer: A new device to test the (un)pleasurableness and tolerance of vaginal pressure and the influence of sexual stimuli. *Journal of Sex & Marital Therapy*, 44(2):189–200.

Previdi, F. (2002). Identification of black-box nonlinear models for lower limb movement control using functional electrical stimulation. *Control Engineering Practice*, 10(1):91–99. Modelling and Control in Biomedical Systems.

Puppo, V. (2011). Embryology and anatomy of the vulva: the female orgasm and women's sexual health. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 154(1):3–8.

Puppo, V. (2013). Anatomy and physiology of the clitoris, vestibular bulbs, and labia minora with a review of the female orgasm and the prevention of female sexual dysfunction. *Clinical Anatomy*, 26(1):134–152.

ter Kuile, M. M., Both, S., and van Lankveld, J. (2010). Cognitive Behavioral Therapy for Sexual Dysfunctions in Women. *Psychiatric Clinics of North America*, 33(3):595–610.

van der Velde, J., Laan, E., and Everaerd, W. (2001). Vaginismus, a Component of a General Defensive Reaction. An Investigation of Pelvic Floor Muscle Activity during Exposure to Emotion-Inducing Film Excerpts in Women with and without Vaginismus. *International Urogynecology Journal*, 12(5):328–331.

Van Lunsen, H. and Ramakers, M. (2002). The hyperactive pelvic floor syndrome (HPFS). *Acta endoscopica*, 32(3):275–285.

Varagnolo, D., Knorn, S., Melles, R., and Dewitte, M. (2017). Qualitative modeling of pressure vs. pain relations in women suffering from dyspareunia. In *Proceedings of the 20th IFAC Worldcongress*.

Varagnolo, D., Knorn, S., Oliver-Chiva, E., Melles, R., and Dewitte, M. (2018). Data-driven modelling of subjective pain/pleasure assessments as responses to vaginal dilation stimuli. *IEEE Control Systems Letters*, 2(3):423–428.