



Model-based selection of most informative diagnostic tests and test parameters

Sven Herrmann and Mathias Dietz* 

Department für Medizinische Physik und Akustik and Cluster of Excellence “Hearing4all”, Universität Oldenburg, Oldenburg 26111, Germany

Received 9 April 2021, Accepted 21 October 2021

Abstract – Given the complexity of most brain and body processes, it is often not possible to relate experimental data from an individual to the underlying subject-specific physiology or pathology. Computer simulations of these processes have been suggested to assist in establishing such a relation. However, the aforementioned complexity and required simulation accuracy impose considerable challenges. To date, the best-case scenario is varying the model parameters to fit previously recorded experimental data. Confidence intervals can be given in the units of the data, but usually not for the model parameters that are the ultimate interest of the diagnosis. We propose a likelihood-based fitting procedure, operating in the model-parameter space and providing confidence intervals for the parameters under diagnosis. The procedure is capable of running parallel to the measurement, and can adaptively set test parameters to the values that are expected to provide the most diagnostic information. Using the pre-defined acceptable confidence interval, the experiment continues until the goal is reached. As an example, the approach was tested with a simplistic three-parameter auditory model and a psychoacoustic binaural tone in a noise-detection experiment. For a given number of trials, the model-based measurement steering provided 80% more information.

Keywords: Hearing, Auditory model, Psychoacoustic, Diagnosis, Measurement procedure

1 Introduction

Audiological diagnostics has the goal of identifying the physiological cause of a hearing impairment, and of quantifying its extent (e.g., [1]). Due to the complexity of the auditory system, it is a far from trivial exercise to identify a specific cause from non-invasively recorded data. Experienced audiologists and elaborate experimental protocols allow for diagnosing most of the common causes, but even in these “easy” cases, the description of the cause is qualitative. The audiologist gathered lots of quantitative data such as an audiogram, a tympanogram, an auditory brainstem response (ABR), and so forth. Inevitably, however, the diagnosis is usually not quantitative with respect to the physiological cause. It is not going to be “patient X has 60–80% loss of auditory nerve fibers at frequencies above 2000 Hz and a 5–10 mV reduced endocochlear potential”. But these physiological parameters would be most useful for providing optimal treatment, at least in theory, if a patient’s auditory system can be fully characterized by a large set of such parameters. To derive physiological parameters from experimental data, an accurate computational model is required.

To date, however, such models only exist for isolated parts of the auditory system; especially models of the

auditory periphery have evolved into comprehensive and broadly-tested frameworks (e.g., [2–4]). A second, equally large, problem is that for diagnostics, we need to calculate the inverse function that brings us from a data set to the model parameters [5]. The number of physiological parameters will be too large to expect by chance to collect a data set that will be large enough to confine the parameter space. These limitations do not, however, need to stop us from developing a theoretical framework for model-based auditory diagnostics. Pioneering work from Panda et al. [6] have already demonstrated the general applicability of models in auditory diagnostics. They used a model of the auditory periphery [2], and altered its parameters to simulate data from several patients. For each patient, they changed only one parameter at a time – all other parameters were kept at their default value. They could then speculate that the respective non-default parameter is a possible description of the respective patients’ peripheral pathology. However, the study was limited to qualitative fitting, and was only used for the analysis of a single, non-default parameter per patient, which runs counter to the common observation of co-morbidities.

With a model at hand, it is also possible to orchestrate the experiment for an efficient diagnosis: It should be possible to calculate which experiment, and which experimental condition, can be expected to generate the most

*Corresponding author: mathias.dietz@uni-oldenburg.de

useful information to confine the parameters. Moreover, with the model steering the experiment, any new data is going to live-update the diagnosis. Parameter estimates and confidence intervals are generated continually and can even be considered to determine the end of the experiment.

A common example is the maximum-likelihood procedure. The maximum-likelihood procedure is an established, adaptive psychophysical measurement technique that sets an adaptive stimulus variable to the value at which most information can be obtained. What “most information” means, can be defined by the operator and will heavily influence the adaptation process. Typically, it is either a threshold value, or the parameters of a psychometric function. Essentially, the psychometric function is already a simple (non-physiological) model that has two or three parameters. A related approach also exists for fast audiogram measurement [7]. Both the classical maximum-likelihood procedure and [7] use a simple “model” of how a stimulation parameter (e.g., level) influences the response. The goal is to minimize confidence intervals on conventional result scales, such as, e.g., a “fraction of correct responses” scale. In contrast, an automated, model-based diagnosis aims at maximizing information about a set of model parameters, e.g., the amount of hair-cell loss. A related approach has been used for a time-efficient estimation of the band importance function parameters of the speech intelligibility index [8] but not for estimating parameters of a processing model.

The main goal of the current study is therefore to provide a first conceptual attempt at model-steered diagnostics. Section 2 describes the concept in very general terms, as it can also be applied to non-auditory diagnostics. In Section 3, the concept will be applied to a specific example. An auditory model will be used to steer a psychoacoustic experiment aimed at identifying three parameters of an artificial test subject.

2 Methods

2.1 Formalism

It is assumed that the behavior of a certain brain or other body function can be described as a function f that has been implemented as a computer model. The goal is to identify the model instance f_j , that is optimal for reproducing experimental responses r of an individual. In this general view, the different instances of f_j can represent one model function operating with different parameter sets, and/or different model functions.

For any experimental condition or stimulus s_i the output of the model $y_{i,j} = f_j(s_i)$ must have the same format as the experimental data $d_i = r(s_i)$; i.e., the model has to operate as an artificial subject. Last, the set of experimental conditions $s_1 \dots s_N$ must allow for disambiguating between the possible model instances. In other words, an s_i must exist, so that any two different model instances, e.g., f_1 and f_2 , produce different outputs.

For solving the inverse problem, it is necessary to compare all model data $y_{i,j}$ against all corresponding

experimental data d_i . The $y_{i,j}$ is effectively a large matrix with simulated data, which can be pre-calculated, if desired. With the introduction of an error cost function $c(d_i, y_{i,j})$, the optimization problem is formalized as a cost minimization:

$$\operatorname{argmin} \left(\sum_{i=1}^N c(d_i, y_{i,j}) \right). \quad (1)$$

2.2 Maximum likelihood

The cost function will typically increase with an increasing difference between the experimental data d_i and the corresponding simulated data $y_{i,j}$. However, it will also critically depend on the probability density functions of d_i and $y_{i,j}$. In theory, the latter can be infinitely precise, but at least the experimental data has some uncertainty. However, it is also possible that the $y_{i,j}$ is a probability distribution rather than a singular value. In any case, when at least one of the two quantities is a distribution, the maximum-likelihood method can be used as an elegant substitute for the general cost function (Eq. (1)). In the following, we assume a precise simulation value $y_{i,j}$ and a probability distribution of d_i . When defining $p(d_i, y_{i,j})$ as the probability of d_i at the simulated value $y_{i,j}$, the likelihood estimator can be written as:

$$L_j = \prod_{i=1}^N p(d_i, y_{i,j}), \quad (2)$$

where L is the estimated likelihood. To be able to derive the function more easily and to avoid numerical problems, the logarithm is applied to the formula:

$$\ln(L_j) = \sum_{i=1}^N \ln(p(d_i, y_{i,j})). \quad (3)$$

The instance j resulting in the maximum of L_j or $\ln(L_j)$ corresponds to the most likely model f_j underlying the data set $d_1 \dots d_N$.

2.3 Parametric approach

Instead of the arbitrary model instances f_j , in most cases a fixed functional relation is assumed, and instances differ only in their parameters $f_j(s_i) = f(m_1, \dots, m_l, s_i)$. The goal is then to determine the model parameters m_1, \dots, m_l that have the maximum likelihood of underlying a certain data set (Fig. 1). While the difference may appear to be only terminological, it quickly becomes essential in practice, when the goal is not only to find the most likely set of parameters, but also to estimate confidence intervals or probability density functions for each model parameter. Analog to the likelihood L_j of instance j , the parametric likelihood is written as $L(m_1, \dots, m_l)$. The total likelihood for parameter m_1 can then be written as:

$$L(m_1) = \sum_{m_2, \dots, m_l} L(m_1, \dots, m_l). \quad (4)$$

In the common case that $L(m_1)$ has a single maximum and tails off monotonically on both sides, diagnostically relevant confidence intervals can be specified. Likelihoods of all other parameters $m_2 \dots m_l$ are derived accordingly.

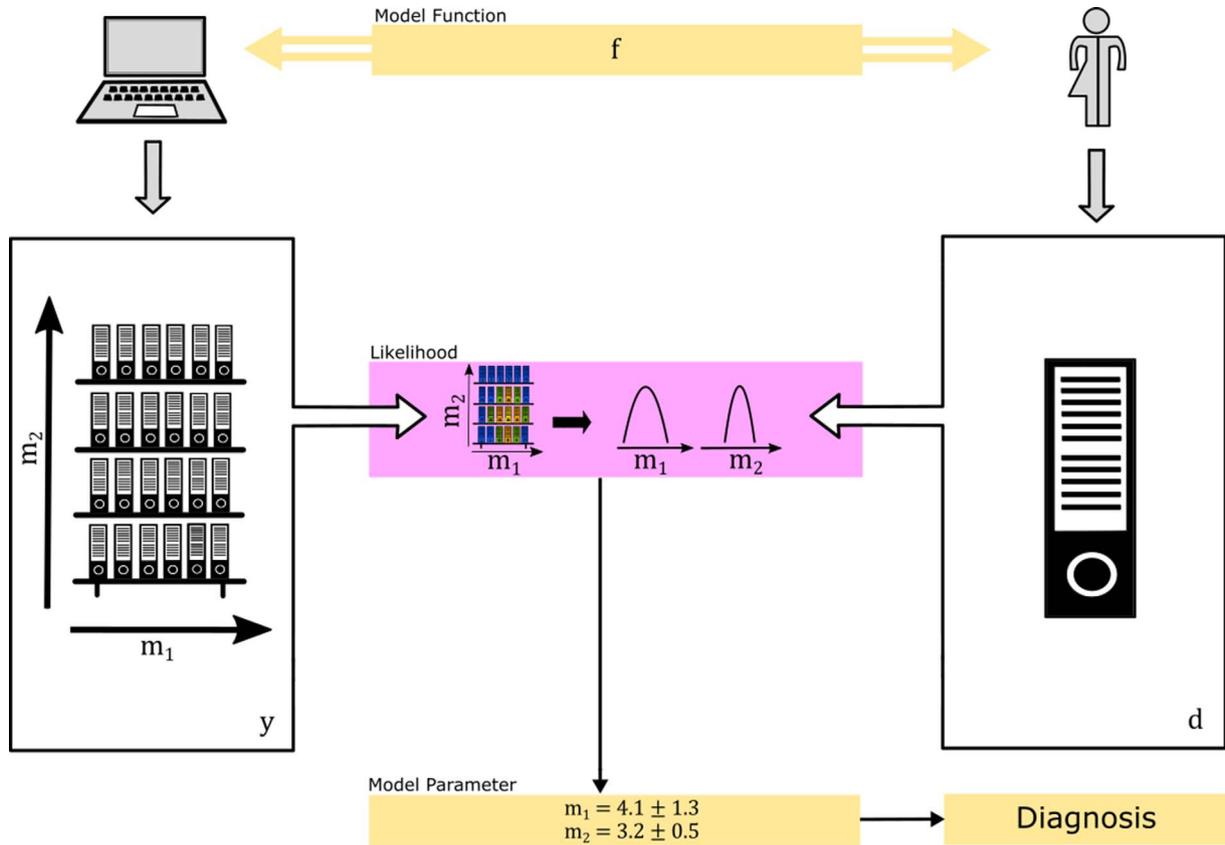


Figure 1. General approach for the model-based diagnostics using a likelihood procedure. The model function is used to generate the look-up table y , and it is assumed that this function works in a same way as the subject. The subject data is obtained experimentally. For illustration, a folder represents all data from either a subject or from a specific model instance. All possible combinations of model parameters make up the folder shelf, representative of the model table y . Consider that for the respective folder, each new measurement generates a new sheet of paper containing the stimulus information and the response. The goal is to find the model instance (folder) that has the maximum likelihood of having generated the experimental data. The resulting likelihood surface (left side of purple rectangular) has the same dimension as the folder shelf. In addition to the likelihood of each instance, confidence intervals can be calculated for each model parameter, which represent the model-based diagnosis.

2.4 Model-based measurement steering

Our diagnostic goal is to minimize or to limit the confidence intervals of the diagnostic parameters. This can be m_1, \dots, m_l or any subset or mathematical function of these primary parameters. For example, using the body-mass-index rather than mass and height.

The skilled experimenter chooses experimental conditions that cause a reduction of the confidence intervals in the diagnostically relevant parameters. In an ideal case, the experiment should be optimized towards efficiently collecting diagnostic information. For this task, the model-based approach is instrumental, but suffers the same problem as the human experimenter: without knowing the parameters of the test subject, it is not possible to define a complete set of experimental conditions upfront. An ongoing adjustment of the experimental conditions at runtime is a solution for this problem.

From a theoretical point of view, it would be ideal to determine the most informative experimental condition,

execute the experiment, update the estimated likelihoods and repeat these steps in a loop until the confidence intervals are sufficiently small or until the allocated time has been expended (Fig. 2). However, this may not be the most practical approach, and in most cases, a cost function for changing the experimental conditions must be added.

The arguably most central question of model-based experiment steering is how to combine the confidence intervals into a single scalar value representing the diagnostic fidelity. In the simplest case, all relative standard deviations can be added, and the sum can be minimized. If some parameters are more important, they can be weighted. If relations of primary parameters are of diagnostic value, the confidence intervals of the respective relations must be used. In most cases, it is not diagnostically optimal to minimize the volume of a multidimensional confidence space. Such a criterion does not lead to small confidence intervals for each parameter, but rather to co-varying parameter estimates, i.e., eccentric (ellipsoid) confidence spaces.

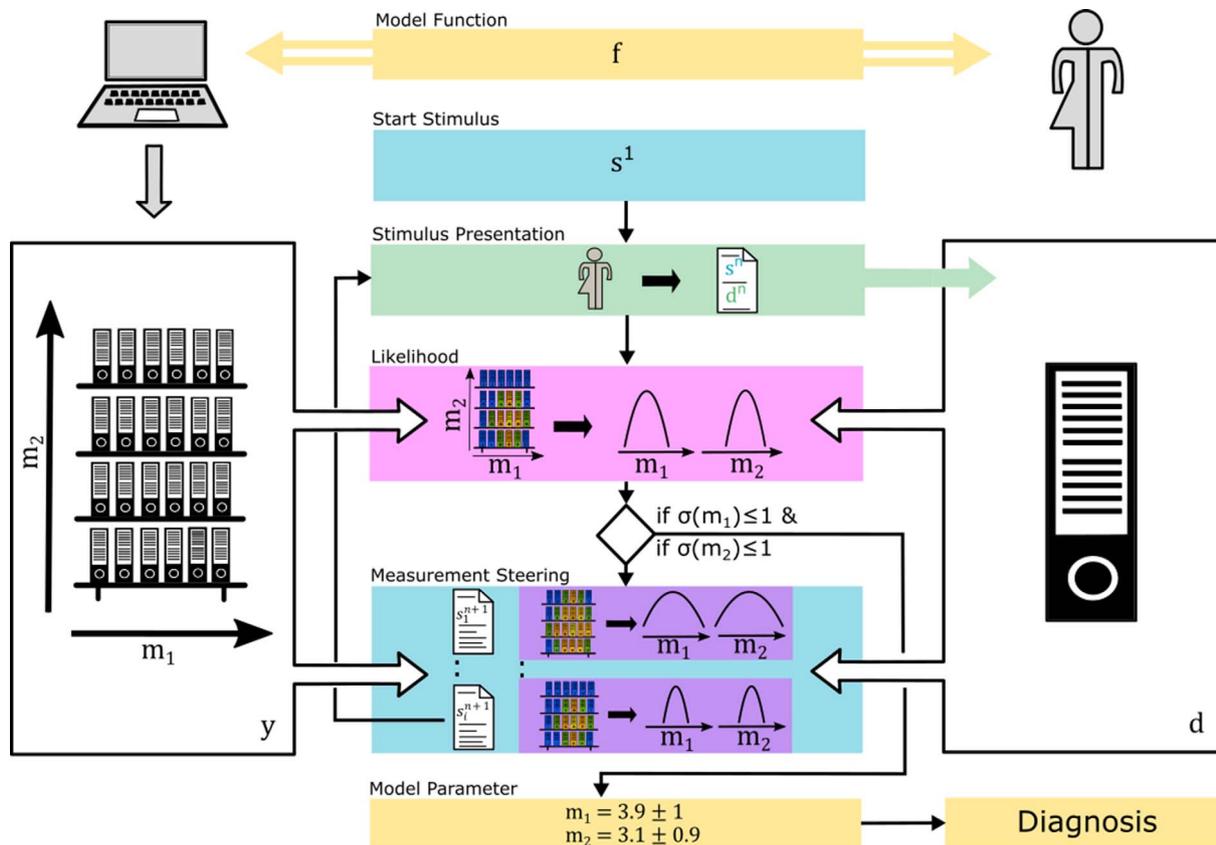


Figure 2. The general approach shown in Figure 1 extended with the model-based measurement steering. The calculation of the likelihood is carried out after every trial, to call up the predefined termination criterion. As long as this is not fulfilled, the next stimulus is calculated, simulating the next trial by adding all possible stimuli s_i to data d . For every simulated stimulus, a likelihood function with the associated standard deviations is also determined in the simulation and the stimulus that leads to the smallest standard deviation of parameter estimates will be presented next.

2.5 Application to psychophysics

So far, the aim was to describe a general concept. Within this subsection, the concept is specified step-by-step until it can be applied to an alternative forced-choice procedure.

First we assume that $y_{i,j} = f_j(s_i)$ is known. If f is an artificial observer, it will provide binary wrong or correct answers, while we require the correct rate. Therefore, many repetitions have to be simulated with the same stimulus until the underlying binomial statistics guarantee the desired accuracy. Furthermore, the large number of different model and stimulus instances cause an enormous computational load, even if a single run of f is very fast. In this case, the simulations have to be carried out before the experiments and $y_{i,j}$ is stored as a look-up table. The concept is much better suited for analytical models that directly provide the correct rate. Many approaches are expected to be between the two extremes, e.g., a numerical model that is capable of estimating the response probabilities more effectively than via an artificial observer.

While model parametrization is quintessential for diagnostic purposes, stimulus parametrization is not necessary within the present concept. The only exception can be a

fundamental parameter such as, e.g., stimulus level, or frequency difference, that is conventionally used as independent, or adaptive variable. It may be necessary to treat each value of such a fundamental variable as a new stimulus instance, but if the psychometric functions have a constant steepness for all model parameter sets and stimuli, different values of the fundamental variable do not need to constitute a new stimulus instance. In this common case, a constant ratio exists between d and the fundamental variable that carries all the information about the psychometric function [9]. The variable should then be set to the estimated most informative value of the psychometric function, as in conventional maximum-likelihood adaptive procedures targeting only the threshold but not the slope [10].

Even beyond such a tentative fundamental variable, the proposed procedure can be understood as closely related to the maximum-likelihood adaptive procedure. The next stimulus is chosen to maximize the information to be learned about the model parameters. The main difference compared to the classic likelihood method is that the fitting function is not a typical sigmoidal psychometric function going from chance to certainty but rather a more abstract model function with more parameters to estimate. Above all, it remains a maximum-likelihood method, intending

to reduce the measurement time. Also, just as in all maximum-likelihood procedures, “unlikely” responses in the early phase of the measurement will increase the duration.

3 Example

In this section, the theoretical concept described in the previous sections is applied to a practical example. Binaural tone-in-noise detection data was simulated for two different stimulus parameters (stimulus bandwidth and noise delay) as well as for tone level, the fundamental stimulus variable. The model has the three parameters auditory filter bandwidth (BW), phase noise (σ_{IPD}), and decision noise (σ_d) [11]. The experimental procedure, the model, and the model-based steering algorithm was implemented in Matlab as a procedure for the psychophysical measurement package AFC [12]. The framework can operate stand-alone, but it can also interface with any experimental software or with any model that provides the basic artificial subject conventions [13]. This also includes several models and data sets from the auditory modeling toolbox [14], for which a general interface is available. The Matlab code is freely available in Zenodo (<https://doi.org/10.5281/zenodo.5211870> [15]).

3.1 Experiment

A 300 ms interaurally antiphase 500 Hz target tone (S_π) was temporally centered in 380 ms of white masking noise with a center frequency of 500 Hz and bandwidths of 25 Hz, 50 Hz, 100 Hz, 200 Hz, 400 Hz, 700 Hz, and 1 kHz. An interaural time difference of 0 ms, 2 ms, 4 ms, 6 ms, 8 ms, or 10 ms was applied to the interaurally correlated noise (N_τ). In addition, an interaurally uncorrelated masking noise condition was used, corresponding to an infinite noise delay. Note that the interaural delay is always an integer multiple of the 500 Hz tone period, resulting in a constant average π phase difference between the target tone and the masking noise. The noise was presented at constant 45.5 dB spectral level. A raised sine gating 20 ms in duration was independently applied to both target and masker after introducing the interaural delay. While the last section referred to stimulus instances that are not necessarily parametrized, in the following, stimuli are described by their two noise parameters: bandwidth and interaural delay (τ).

The measurement was set up as a 3-interval, 3-alternative forced-choice (AFC) procedure. Two reference intervals contained only the masker, whereas the randomly chosen target interval contained the masker with the tone.

3.2 Model

The model used in this example is a slightly modified version of the so-called IPD model [16, 17]. The auditory filters are simulated by a gammatone filterbank [14, 18]. Here, only a single 4th-order filter with the center frequency equal to the 500 Hz target was employed. In contrast to [16, 17], the bandwidth of the filter was not set to a fixed value, but constituted the first diagnosis parameter, e.g., to phe-

nomenologically model the consequences of an outer hair-cell loss (e.g., [19]). Very simplified hair-cell processing is simulated by half-wave rectification and a subsequent 5th-order low-pass filter with a cut-off frequency of 770 Hz [20]. In addition, the signal was taken to the power of 0.4 to simulate compression. Binaural processing was then simulated by an IPD extraction process with high temporal resolution (for details see [17]). White noise was added to the resulting phase difference, to simulate binaural processing limitations. The standard deviation of this noise was adjustable and constituted the second model parameter.

The model concept is based on the assumption that the tone is detectable in the noise masker if it induces a large enough increase in IPD fluctuations. Therefore, the decision stage first quantifies these IPD fluctuations by calculating their mean deviation from zero. After taking the cosine of the IPD fluctuation, the obtained value is mostly identical to the classical cross-correlation coefficient [21]. To map changes of this value to a scale that is proportional to a subjects’ correlation discrimination sensitivity, a Fisher Z -transformation is applied. The individual discrimination sensitivity on this scale is simulated with the third diagnostic parameter: A random value is taken from a Gaussian distribution with a standard deviation of σ_d and added to the resulting fisher’s Z -transformed quantity to form the final decision variable. This is simulating an imperfect internal representation at the level of the decision-making process. Last, the artificial observer expects the tone to be present in the interval with the lowest decision variable.

3.3 Conventional simulation

To get a better understanding of the effect of the individual model parameters on the detection thresholds, the corresponding experiment was simulated with different parameter settings of the model using a 3-interval, 1-up 2-down procedure [22]. Initially, the level was changed by 4 dB and, after the 2nd and 4th reversal, the step size was reduced to 2 and 1 dB, respectively. An additional 6 reversals were measured at this step size and the average of these reversal levels was taken as the threshold obtained in that run. In each condition, 10 runs were simulated. The simulation results obtained with four different model parameters sets are shown as an example in Figure 3. The first set resembles the assumed parameters of a well-performing, normal-hearing subject (Fig. 3A). The values selected here were taken from Dietz et al. [11], but also agree well with data from Bernstein and Trahiotis [23]. In each of the other sets, one parameter has been doubled while the other two remain at the default value. All three model parameters affect the simulated thresholds of the experiment in different ways. This is an important prerequisite for the inverse problem of model-based diagnostics: to identify the model parameters from the data.

Specifically, an increase in filter bandwidth increases thresholds for broadband noise maskers, while leaving 25- and 50 Hz noise thresholds unchanged (Fig. 3B). An increase in phase noise primarily effects the conditions with zero or small values of tau (Fig. 3C), whereas the decision

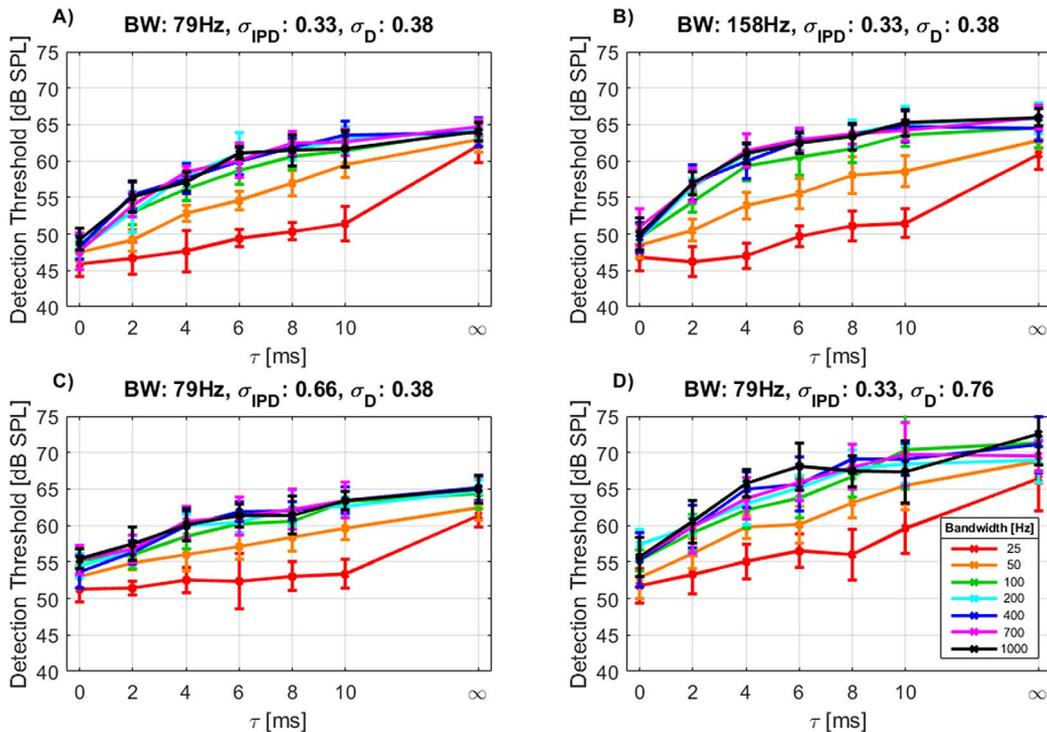


Figure 3. Simulated detection thresholds with different model parameter combinations using the Dietz model [10, 13]. BW: bandwidth of the auditory filter, σ_{IPD} : standard deviation for the phase noise, σ_{D} : standard deviation for the decision noise.

noise increases the thresholds for all conditions fairly uniformly, but also increases the variances (Fig. 3D).

3.4 Model-steered simulation

To pre-calculate the required look-up table y , psychometric functions for all 49 stimuli were simulated using all possible combinations of model parameters. Each model parameter range was discretized into 8 values with a 1/3-octave spacing, 0.2–1.0 for the two noise parameters, and 40–200 Hz for the bandwidth, resulting in 512 possible model parameter sets and a total of 25 088 test conditions. In order to obtain the desired correct rate for each of these conditions, the binary simulation result (correct or wrong interval) requires repetitive testing to obtain a correct rate or even the psychometric function. As deduced in Section 2.5, the correct rate at a single level is sufficient in the case of a condition-independent psychometric function steepness [9]. In the present example, however, the steepness changes somewhat with σ_d . We therefore decided to store both steepness and a threshold value for each condition. To derive the two parameters, a psychometric function was estimated from 1500 stimulus presentations distributed equally from 25 to 75 dB in a 1.5 dB grid and a subsequent logistic fit. Arguably more elegantly, the steepness and slope can also be obtained following the classical maximum-likelihood procedure [24]. The threshold was defined at the most informative level of the psychometric function [25], also called the sweet point. In the current example, this point is at a correct rate of 72.9%.

Once the table was generated the actual model-steered measurement as described in Section 2.4 started: After each response, the likelihood was calculated in the 3-dimensional parameter space (Sect. 2.2). From this likelihood space, the three standard deviations were calculated (Sect. 2.2) and summed to obtain the final variable to be minimized. Moreover, this variable was also calculated for all 49 possible next stimuli. For each stimulus, the updated variable was calculated for both an assumed correct and an assumed wrong response from the subject. In order to obtain a single value from each of the 49 pairs, the two values of each pair were averaged, but weighted according to their expected probability of occurrence (i.e., 72.9% correct, 27.1% wrong). Then the stimulus with the smallest weighted average was selected for the next presentation. This was repeated until the previously defined criterion was met. In our case, the average relative standard deviation of the three parameters was below 10%, or 1500 trials had been completed.

3.5 Results and discussion

Figure 4 shows an example result of the procedure as described in Section 3.4 with an artificial test subject. The parameter set for this artificial test subject correspond to that of a normal-hearing person [11]: BW: 79 Hz, σ_{ITD} : 0.33, σ_{D} : 0.38. The values were deliberately chosen so that they are not contained directly in the model table y , but lie between the values in the table. The stimulus parameter set for the first trial was arbitrarily set to $\tau = 0$ ms and a bandwidth of 25 Hz at 65 dB SPL. At this high level, a “correct”

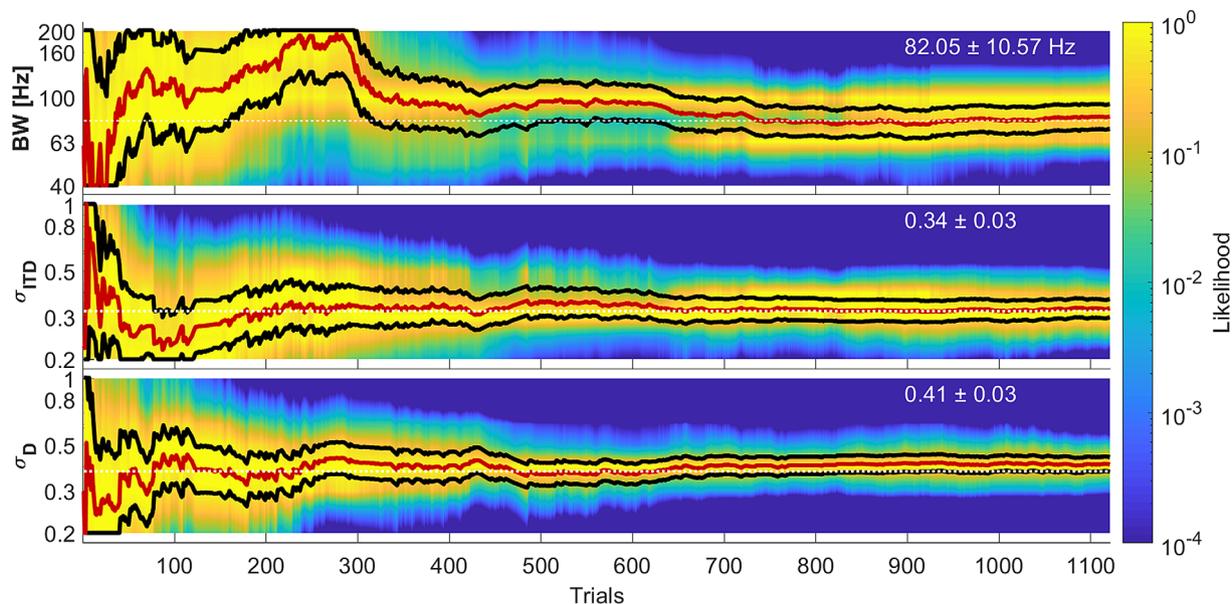


Figure 4. Example result of the model-based measurement steering with an artificial test subject. In addition to the likelihood (colored), the estimated mean (red line) and the $1 \times \sigma$ confidence intervals (black lines) are shown. The white line represents the true internal values (BW: 79 Hz, σ_{ITD} : 0.33, σ_D : 0.38). The last estimates of the model parameter are noted as text in the corresponding panel.

response can be expected for all (artificial) subjects, a common choice for maximum-likelihood procedures.

After about 500 trials, the phase noise and decision noise could be estimated with an accuracy similar to the discretization step size. In the displayed example, the first termination criterion, a relative standard deviation of only 10%, was met after 1121 trials. Note that this standard deviation is even less than the discretization step size. The slightly larger relative standard deviation of the bandwidth estimate can be explained by the fact that this parameter – compared to the other two parameters – in general has a smaller influence on thresholds. A single step (i.e., 33%) change in estimated phase noise (as it happened near trial 800) caused an immediate 67% change in estimated filter bandwidth, because the two parameters are covariant, but with different strong impacts at this point of the parameter space.

As we used an equal weighting of all three diagnosis parameters, it was very important that all parameters had a similar lever on the thresholds, to minimize problems such as that just described. If a discrete parameter space is used, the threshold change caused by a single model parameter step should be of the same order of magnitude as behaviorally discriminable sensitivity levels, i.e., a few dB in the present example. Across all possible stimuli, changing a single model parameter by one step (here 33%) resulted in a change of simulated detection thresholds of up to 4.3 dB for phase noise (almost only for stimuli with $\tau = 0$), up to 1.9 dB for a step change in decision noise, but similar for all stimuli, and up to 3.6 dB for bandwidth, but only for wider-band stimuli with $\tau = 2$ ms or $\tau = 4$ ms. While grid optimization may further improve performance, the present

grid appears to offer three fairly comparable dimensions with reasonable discretization steps.

It was also noticeable that the mean from the fitted Gaussian can deviate from the maximum of the likelihood. This is because of the occasionally asymmetric shape of the likelihood. Therefore, the Gaussian fit is problematic, but still has practical merit.

Overall, after 1121 trials, the estimated filter bandwidth (true value 79 Hz) was estimated to be 82.05 ± 10.57 Hz, the phase noise (true value 0.33) was estimated to be 0.34 ± 0.03 and the decision noise (true value 0.38) was estimated as 0.41 ± 0.03 .

In [Figure 5](#), the same experiment was carried out as above ([Fig. 4](#)), with the difference that the model used as an artificial test subject was modified: instead of the gammatone filter bank used to simulate the basilar membrane, a DRNL (dual resonance nonlinear filter) filter bank was used [26]. Therefore, the model used to calculate the model table η was no longer identical to the model that functioned as an artificial test subject. Visual inspection of [Figure 5](#) and a comparison of the estimated parameters reveal that despite the modification, the system was able to meaningfully estimate the parameters of the test subject.

As in [Figure 4](#), this initially only applied to the combination of parameters that represent a normal-hearing subject. The next step was to check the extent to which the system is able to correctly estimate different parameter combinations, and thus subjects with different hearing abilities. [Figure 6](#) shows the final estimates of the internal parameters of different artificial test subjects with different parameter combinations: Four different values were selected for each of the three parameters and simulations

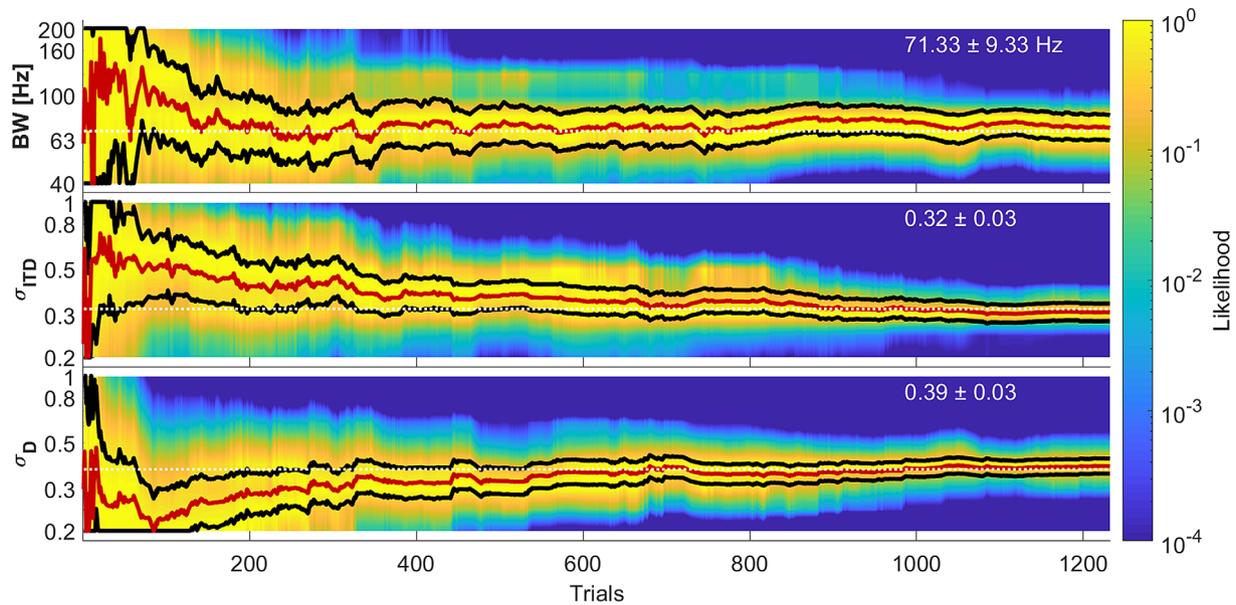


Figure 5. Example result of the model-based measurement steering with an artificial test subject. In general, the same experiment is shown as in Figure 4 with the difference that here, the model used as an artificial test subject was modified: the gammatone filter, which simulates the basilar membrane, was replaced with a DRNL filter. Due to this new filter, the equivalent rectangular bandwidth was 69 Hz (at 70 db SPL), instead of 79 Hz in the previous example. The other two parameters were identical to the previous example (σ_{ITD} : 0.33, σ_D : 0.38).

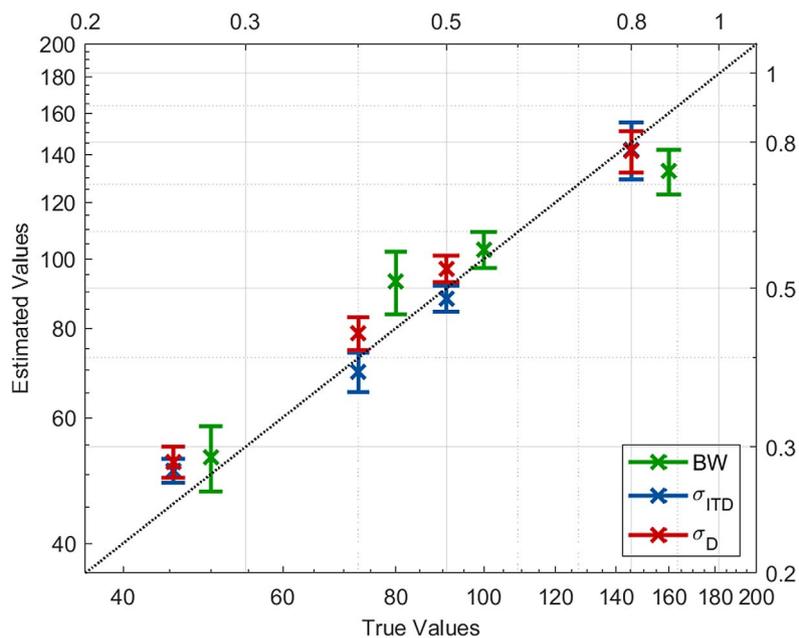


Figure 6. Final estimates of the internal parameters of the artificial test subject using 64 different parameter combinations. The results of the individual parameters are presented using a 95% confidence interval. Note that the resolutions of the 3 parameter values are the same, but to obtain better readability, the abscissa for the bandwidth is shown in an abbreviated form.

were carried out for all of the resulting 64 parameter combinations. Each simulation was stopped after the accuracy for all three parameters reached the $1 \times \sigma$ confidence interval. The model chosen for the artificial test subject contained the gammatone filter bank. It can be seen that the two noise parameters can be estimated relatively accu-

rately, regardless of their value, although the accuracy decreased somewhat at high values. The bandwidth estimate was also generally good, but with slightly larger confidence intervals. One exception was the underestimation of filter bandwidth at 160 Hz. Due to the above-mentioned covariance of bandwidth and phase noise, the bandwidth

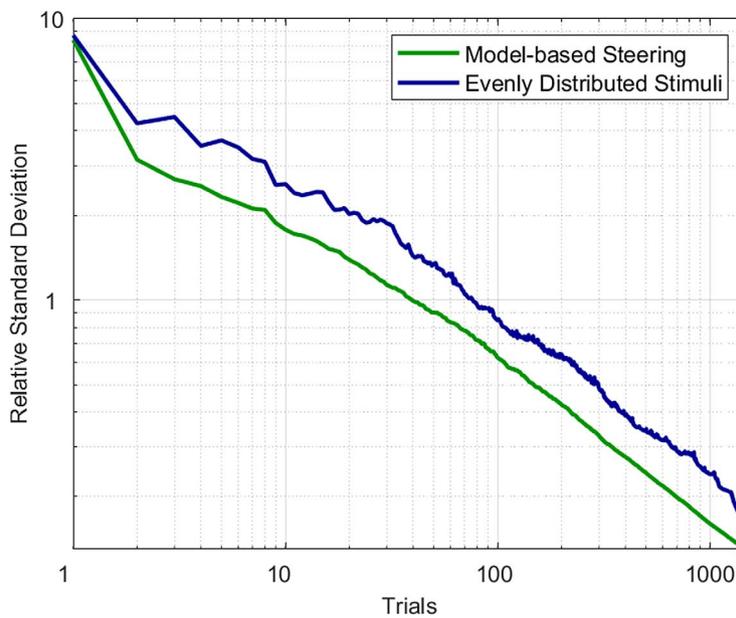


Figure 7. Mean relative standard deviation of the three model parameters as a function of trial numbers.

likelihood function was not always Gaussian, but sometimes had noticeable side peaks (see Fig. 4), each corresponding to a certain phase-noise value. In addition, the bandwidth likelihood functions were arbitrarily limited by our grid to 200 Hz, truncating the upper side peak for the 160-Hz condition. Hence, the fitted mean value was systematically biased towards lower values.

With conventional maximum likelihood, but without the model-based steering, the relative standard deviations were about 1.3–1.5 times larger (see Fig. 7). Equivalently, without the model-based steering, almost 80% more trials (e.g., for an average standard deviation of 33%) were required to achieve the same accuracy.

Insights can also be obtained by analyzing which stimuli were selected by the steering algorithm, and perhaps even in which temporal order. In the present example, it was observed that the algorithm primarily selected stimuli from the edges of the possible parameter ranges, and from 2 to 4 ms delays at large bandwidths (Fig. 8A).

Within the first 20 trials, the stimulus selection was fairly equally distributed, with a small preference for 25 Hz (Fig. 8B). A little later in each run, the 25 Hz preference is more pronounced. This behavior can be understood by comparing to Figure 3: At first, the only parameter that could be estimated irrespective of the other two parameters is decision noise. All conditions were influenced by decision noise in a similar way, but especially the 25 Hz conditions were not influenced by filter bandwidth at all, so they are initially overrepresented. Particularly, the 25 Hz, $\tau = \infty$ condition was not influenced by either bandwidth or phase noise, so it is, at least in retrospect, understandably the most frequent condition and always caused an improvement of the estimates of decision noise. In Figure 3, the largest influence of filter bandwidth was observed under conditions with a large noise bandwidth and $\tau = 2$ or 4

ms. These conditions were also moderately overrepresented (see Fig. 8).

4 General discussion

We have demonstrated conceptually (Sect. 2) and via simulation examples (Sect. 3) that a processing model of a brain- or body function can be used to steer a measurement procedure such that the selected measurement instance maximizes the information about the respective function. In a diagnostic context, “information” refers to model parameters, and the information maximization process to the minimization of confidence intervals. Within a research context, the same approach may also help to select measurement conditions that optimally disambiguate between different model concepts, i.e., different functional I/O relations, irrespective of parameters. From a user perspective, in addition to the time efficiency, the main benefit is that the diagnostic accuracy is inherently tracked. This allows for well-informed termination criteria, and avoids negative surprises about missed levels of significance that can happen with the common sequential measurement – analysis approach.

A critical element of the proposed framework is the optimization criterion. In the present example, the sum of three standard deviations was chosen to be minimized. If diagnostic interest exists only in some parameters, the other parameters can be left out of the optimization criterion, but will still be confined to some degree as a by-product. Further, a diagnostic priority can be incorporated as an increased weighting of the respective parameter’s confidence interval. A preliminary version of the framework used the volume of the 3-dimensional likelihood space that exceeded a certain likelihood threshold. This method was of no diagnostic

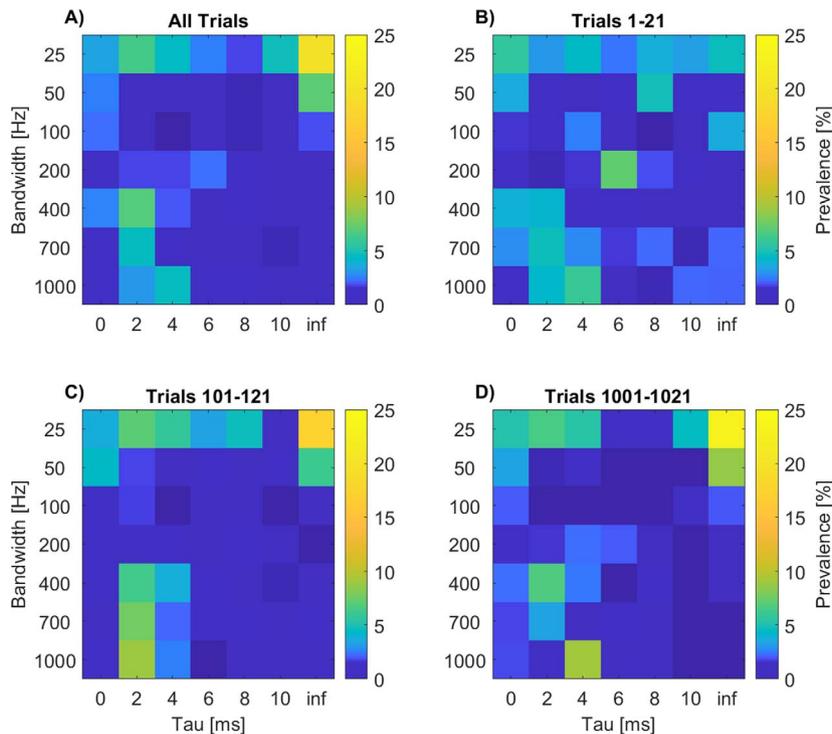


Figure 8. Prevalence of the stimuli selected by the model-based steering as percentages. Once over all trials (A) and, additionally, in certain parts of the experiment (B–D). The data shown are averaged over 100 experiment runs, each of which consisted of 1470 trials.

value, because there was no incentive to disentangle co-varying parameters. In fact, such a simple approach may try to maximize co-variances. Illustratively speaking, in a 2-dimensional space, our approach is operating towards a circle-shaped confidence area, whereas the preliminary version resulted in strongly ellipsoid areas, or lines as confidence areas: a smaller area (or volume in general), but no parameter confinement. The optimization criterion also has to match the diagnostic requirements. In the present example (Sect. 3), it can happen that two parameters are very accurate and one has a larger confidence interval. If the diagnosis requires a maximum confidence interval for certain parameters, the optimization criterion also has to be modified, to ensure that the procedure targets the termination criterion. Last, in the case of a discrete implementation, the optimization process can depend on the resolution of the model parameter grid. In the process of designing the example, a discretization step size that caused notable differences in simulated results, in at least one condition, was found to be ideal. Again, to optimize this critical design parameter, diagnostic ambition and measurement efficiency must be aligned. If effect sizes of steps of the various parameters differ, additional problems may occur. For example, a single discrete step in one dimension may then cause a multi-step covariance of a second parameter, which can create severe problems for the confinement of the second parameter. Near the boundary of the parameter grid, such a covariance can also cause a bias, as observed in one condition in Figure 6.

Despite the proof of concept, various obstacles currently prohibit practical applications. First and foremost, an accurate and detailed general-purpose model is required that captures the complete I/O chain from stimulation to recording or perception. Any simplification (which is inevitable) causes the risk of hidden parameters that may corrupt the diagnosis. Even if we assume that a general-purpose model will exist at some point in the future, it will likely have far too many parameters to be able to apply the proposed procedure. At least this is already the case for today’s most comprehensive auditory models [3, 4]. The only chance appears to be in a very short processing chain, e.g., direct measurements of ear processes, such as tympanometry [27], or otoacoustic emissions. Already the model of a single neuron and its presynaptic input cannot be unambiguously parametrized, even with a multidimensional stimulus parameter space [28]. However, any diagnosis is always a quantitatively imperfect estimate, and must rather be evaluated based on its practical merits. If several parameters or several models have similar consequences and cannot be easily disambiguated, it may not be meaningful to increase the effort until they are disambiguated. The most likely way forward is through highly simplified models that may need to be specifically developed for such a purpose. The two parameters of the Plomp model [29], for instance, are of outstanding practical importance [30], even though they are only very indirectly related to physiology.

Schuknecht [31, 32] presented a very promising dataset from human inner ears that showed pathological causes of

hearing loss. A distinction was made between four predominantly pathological types: sensory, neural, metabolic, and mechanical. This study showed the relationship between these four different types and their different effects on the audiometric hearing threshold. Since the various pathologies can be mapped with the aid of just a few parameters, and this was brought into correlation with a very common measurement, this resulted in a high diagnostic potential. The different phenotypes could also be used successfully for different machine-learning classifiers to determine the etiology of a hearing loss [33].

Another practical problem is the – potentially very large – changes of stimulus conditions or even a frequent change of experiment or task. Normally, behavioral experiments have carefully designed orders in the presentation, to counterbalance training effects, or to allow for familiarization with certain tasks or stimulus features. To date, such aspects are usually not included in models, so that the variable measurement conditions may not only be at times irritating, but a possible reduced subject performance will be misinterpreted by the model. A potential solution could be a blocked presentation, in which the steering process is only performed after a given number of constant stimuli is presented. A second option is to add a parameter change cost-function that reduces the model's willingness to jump within the parameter space (Sect. 2.4). Nonetheless, attention, motivation, training, fatigue, and other similar influences can be expected to remain as a significant limitation of the suggested approach when used in the context of psychophysical experiments.

Yet a further problem arises from the computational load of most models. First, the calculation of the multidimensional model table may require dedicated computational resources. Second, the computation of the steering procedure has to be completed while the subject is given the previous stimulus, otherwise a time-lag to the next presentation will reduce the duration benefit, and will be potentially annoying for the subject.

Critically speaking, the diagnosis of the artificial observer (Sect. 3) is to model a model with itself. Getting out what was put in is the lowest possible target. The minor variation of training model and testing model demonstrated in Figure 5 is a first step towards larger discrepancies between any biological system and its model. However, even such a self-validation has the advantage that the steering process can be studied without being concerned with unaccounted influences. For example, if the actual measurements are to be conducted in a conventional way, a prior investigation of the steering with differently parameterized artificial observers can reveal very useful insights about how to design the measurement. At the very least, it informs which test conditions are the most informative and is thus instrumental in defining the measurement protocol. We had not expected to find that the uncorrelated 25-Hz wide noise masker was by far the most informative (Fig. 8) - but after studying the steering process, it appears to be obvious. Moreover, different parameters can be used to test whether certain artificial subjects require very different test conditions than others. If this is the case, a decision-tree-based

design may help to select certain testing conditions only for some patients. This would constitute a hybrid approach between conventional measurements and the proposed technique, in a similar spirit to work by Sanchez Lopez et al. [30, 34].

5 Conclusion

We conclude that model-based experiment steering is possible and has at least theoretical advantages over sequential measure-and-fit approaches. Practical problems and the lack of sufficiently accurate models will initially prohibit most diagnostic applications. The method will rather be instrumental in providing insights for designing better condition tables and experimental decision trees – but experiments are expected to be mostly executed in a conventional way.

Acknowledgments

We thank Anna Dietze for helping to transfer the code into the AFC-framework and for very helpful suggestions regarding the graphical representations and general feedback in the system. This work was supported by the European Research Council (ERC) under European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 716800 (ERC Starting Grant to Mathias Dietz).

Data availability statement

The code for the model-based-diagnostic procedure, including AFC and scripts for plotting the figures, is available in Zenodo: <https://doi.org/10.5281/zenodo.5211870> [15].

References

1. S. Hoth, I. Baljić: Current audiological diagnostics. *GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery* 16 (2017) Doc09.
2. R. Meddis: Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *The Journal of the Acoustical Society of America* 119, 1 (2006) 406–417.
3. I.C. Bruce, Y. Erfani, M.S.A. Zilany: A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing Research* 360 (2018) 40–54.
4. S. Verhulst, A. Altoè, V. Vasilkov: Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing Research* 360 (2018) 55–75.
5. S. Zenker, J. Rubin, G. Clermont: From inverse problems in mathematical physiology to quantitative differential diagnoses. *PLoS Computational Biology* 3, 11 (2007) e204.
6. M.R. Panda, W. Lecluyse, C.M. Tan, T. Jürgens, R. Meddis: Hearing dummies: Individualized computer models of hearing impairment. *International Journal of Audiology* 53, 10 (2014) 699–709.

7. X.D. Song, B.M. Wallace, J.R. Gardner, N.M. Ledbetter, K. Q. Weinberger, D.L. Barbour: Fast, continuous audiogram estimation using machine learning. *Ear and Hearing* 36, 6 (2015) e326.
8. Y. Shen, A.J. Kern: An Analysis of Individual Differences in Recognizing Monosyllabic Words Under the Speech Intelligibility Index Framework. *Trends in Hearing* 22 (2018) 2331216518761773.
9. H. Dai, C. Micheyl: Psychometric functions for pure-tone frequency discrimination. *The Journal of the Acoustical Society of America* 130, 1 (2011) 263–272.
10. D.M. Green: A maximum-likelihood method for estimating thresholds in a yes–no task. *The Journal of the Acoustical Society of America* 93, 4 (1993) 2096–2105.
11. M. Dietz, J. Encke, K. Bracklo, S.D. Ewert: Prediction of tone detection thresholds in interaurally delayed noise based on interaural phase difference fluctuations. arXiv preprint: [arXiv:2107.00320](https://arxiv.org/abs/2107.00320) (2021).
12. S.D. Ewert: AFC – A modular framework for running psychoacoustic experiments and computational perception models, in *Proceedings of the International Conference on Acoustics AIA-DAGA*, 2013.
13. M. Dietz, J.H. Lestang, P. Majdak, R.M. Stern, T. Marquardt, S.D. Ewert, W.M. Hartmann, D.F.M. Goodman: A framework for testing and comparing binaural models. *Hearing Research* 360 (2018) 92–106.
14. P.L. Søndergaard, P. Majdak: The auditory modeling toolbox, in *The technology of binaural listening*. Springer, Berlin, Heidelberg. 2013, pp. 33–56.
15. S. Herrmann, M. Dietz: Matlab Code for Model-based selection of most informative diagnostic tests and test parameters [Online]. Available at: <https://doi.org/10.5281/zenodo.5211870> [Accessed: Nov 24 2021]
16. M. Dietz, S.D. Ewert, V. Hohmann, B. Kollmeier: Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences. *Brain Research* 1220 (2008) 234–245.
17. M. Dietz, S.D. Ewert, Volker Hohmann: Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication* 53, 5 (2011) 592–605.
18. V. Hohmann: Frequency analysis and synthesis using a Gammatone filterbank. *Acta Acustica United with Acustica* 88, 3 (2002) 433–442.
19. B.R. Glasberg, B.C.J. Moore: Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America* 79, 4 (1986) 1020–1033.
20. J. Breebaart, S. Van De Par, A. Kohlrausch: Binaural processing model based on contralateral inhibition. I. Model structure. *The Journal of the Acoustical Society of America* 110, 2 (2001) 1074–1088.
21. L.R. Bernstein, C. Trahiotis: The normalized correlation: Accounting for binaural detection across center frequency. *The Journal of the Acoustical Society of America* 100, 6 (1996) 3774–3784.
22. H. Levitt: Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America* 49, 2B (1971) 467–477.
23. L.R. Bernstein, C. Trahiotis: Binaural detection as a joint function of masker bandwidth, masker interaural correlation, and interaural time delay: Empirical data and modeling. *The Journal of the Acoustical Society of America* 148, 6 (2020) 3481–3488.
24. M.M. Taylor, C.D. Creelman: PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America* 41, 4A (1967) 782–787.
25. M.R. Leek: Adaptive procedures in psychophysical research. *Perception & Psychophysics* 63, 8 (2001) 1279–1292.
26. E.A. Lopez-Poveda, R. Meddis: A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America* 110, 6 (2001) 3107–3118.
27. B. Sackmann, E. Dalhoff, M. Lauxmann: Model-based hearing diagnostics based on wideband tympanometry measurements utilizing fuzzy arithmetic. *Hearing Research* 378 (2019) 126–138.
28. M. Dietz, L. Wang, D. Greenberg, D. McAlpine: Sensitivity to interaural time differences conveyed in the stimulus envelope: Estimating inputs of binaural neurons through the temporal analysis of spike trains. *Journal of the Association for Research in Otolaryngology* 17, 4 (2016) 313–330.
29. R. Plomp: Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America* 63, 2 (1978) 533–549.
30. R. Sanchez Lopez, F. Bianchi, M. Fereczkowski, S. Santurette, T. Dau: Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends in Hearing* 22 (2018) 2331216518807400.
31. H.F. Schuknecht, K. Watanuki, T. Takahashi, A.A. Belal Jr, R.S. Kimura, D.D. Jones, C.Y. Ota: Atrophy of the stria vascularis, a common cause for hearing loss. *The Laryngoscope* 84, 10 (1974) 1777–1821.
32. H.F. Schuknecht, M.R. Gacek: Cochlear pathology in presbycusis. *Annals of Otology, Rhinology & Laryngology* 102, 1 suppl (1993) 1–16.
33. J.R. Dubno, M.A. Eckert, F.-S. Lee, L.J. Matthews, R.A. Schmied: Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology* 14, 5 (2013) 687–701.
34. R. Sanchez-Lopez, M. Fereczkowski, T. Neher, S. Santurette, T. Dau: Robust data-driven auditory profiling towards precision audiology. *Trends in Hearing* 24 (2020) 2331216520973539.

Cite this article as: Herrmann S. & Dietz M. 2021. Model-based selection of most informative diagnostic tests and test parameters. *Acta Acustica*, 5, 51.