Auditory model-based selection of the most informative experimental conditions

Anna Dietze[1,2,*], Anna-Lena Reinsch[1,2], Jörg Encke[1,2], Mathias Dietz[1,2]

1 Department für Medizinische Physik und Akustik, Universität Oldenburg
2 Cluster of Excellence "Hearing4all", Oldenburg
** Contact author: anna.dietze@uni-oldenburg.de*

Abstract

Identifying the causes of a person's hearing impairment is a challenging task. Even though a broad range of measurement techniques exist, links between the results of one or several listening tests and possible pathologies need to be found. Drawing conclusions from measurement results that were influenced by pathologies in this highly non-linear auditory system remains very difficult. In addition, measurement time is restricted, especially in clinical settings. A central but difficult goal is to maximize the diagnostic information that is collectable within a certain time frame. Computer models simulating auditory processing and possible impairments could be employed to assist in such diagnostics. By using the model-based experiment-steering approach introduced in Hermann and Dietz (2021, Acta Acustica, 5:51), the current study demonstrates its applicability using five young, normal-hearing subjects. In the model-based selection procedure, those stimuli providing the most information about the model parameters were identified in parallel to the measurement, and subsequently presented to the participant. The same binaural tone-in-noise detection task was conducted with two measurement procedures: A standard adaptive staircase procedure and the model-based selection procedure. For this proof of concept, an existing auditory processing model was adopted. Its four free parameters enabled the characterization of the subjects' 250 Hz channel. The model parameters best predicting the subject's sensitivity to a diotic and various dichotic conditions, were obtained using a maximum-likelihood approach. On average, the same accuracy of model parameter estimation was reached 2.5 times faster with the model-steered procedure compared to the standard adaptive procedure. Difficulties regarding the choice of a reliable model and issues to be considered when deciding on reasonable discretization steps of the model parameters are discussed. Although the physiological causes of an individual's results cannot be diagnosed with this procedure, a characterization in terms of functional parameters is possible.

Introduction

The aim of audiological diagnostics is to identify the causes of a person's hearing impairment. A broad range of measurement techniques covering all kinds of deficits in the auditory system is available (for a review see Hoth & Baljic, 2017). To achieve a good diagnosis, comprehensive test batteries including subjective and objective tests are usually carried out as a first step. While some measurements are specific to test for a particular pathology, often a combination of tests is required to differentiate between different causes. This linking of data to the underlying cause or pathology is the second step of the diagnostic process, and poses challenges for audiologists, ENT doctors, and researchers alike, for two main reasons. First, a variety of pathologies and their combinations can cause a similar outcome. Therefore, the realization that more data on a particular experiment or stimulus would have been required often comes subsequent to the data collection. Obtaining this data is sometimes no longer practically possible and often inconvenient. Even if data would exist in abundance, a second challenge remains: The auditory system consists of several highly non-linear stages intertwined with multiple efferent regulations. An experienced professional might be able to interpret the data and relate it to a unique pathology, but such diagnosis remains qualitative. A quantitative description of pathology-descriptive parameters with confidence ranges could provide information such as: The estimated loss of type I auditory fiber synapses is 25%, and ranges between 20% and 30%.

Computer models have been suggested as possibly assistants in relating data to potential pathologies. Panda et al. (2014) used a physiological model of the cochlea (Meddis, 2006) to simulate data from a psychoacoustic test battery from hearing-impaired listeners. By varying one model parameter at a time, they created individualized computer models that enabled suggestions on underlying pathologies of their patients, although a combination of parameters would have yielded even better results in some cases.

Comprehensive physiological models of the auditory system require a large number of parameters to be confined (e.g. Verhulst et al., 2018). In addition, physiological redundancies and co-dependencies in the system are useful to stabilize auditory perception against small disturbances or minor impairments, but they also lead to ambiguities in confining model parameters (e.g. Klug et al., 2020). Functional models, on the other hand, require fewer, though more abstract, parameters, such as filter bandwidth, internal noise, or attenuation. For instance, Plomp (1978) presented a quantitative model predicting speech understanding in noise that had only the two parameters attenuation and distortion. Confining these parameters does not lead to a description in terms of physiological characteristics. Nevertheless, such functional models can help with profiling hearing impaired persons and can predict the benefit to be expected from a hearing aid or hearing prosthesis.

The amount of experimental data required to confine the model parameters depends critically on two factors: Measurement accuracy and – most of all – the number of free model parameters. A single parameter can often be estimated from data obtained within a few minutes (e.g. Brand & Kollmeier, 2002). Appraisal of three parameters, however, can already be expected to require several hours of data collection, at least in psychophysics (e.g., Herrmann & Dietz, 2021). In many cases, it may be prudent to adjust the measurement, based on interim results. The approach of Sanchez Lopez et al. (2018) for instance, can identify the most informative predictors in an auditory test battery, based on the preceding results. Instead of conducting all tests on each individual, only a subset of tests is sufficient for the characterization of listeners. These tests represent the nodes of a decision tree that lead to different diagnoses. Another way to confine the assessment of model parameters in a theoretically most time-efficient way is a maximum likelihood-based procedure running in parallel to the measurement, and selecting those stimuli or tests that cause the best refinement in model parameters (Herrmann & Dietz, 2021). So far, this approach has only been tested with a simulated patient. Theoretically, it can be used with any model and experiment. Nevertheless, the demands on the chosen model are high. It must provide good fits to all data without too many parameters. Otherwise, systematic deviations between model and data under any one experimental condition may cause the procedure to overemphasize this condition or to cause some other form of undesired behavior. Also, co-dependencies of the model parameters should be at a minimum.

The goal of the present study was to test the feasibility of model-based experiment steering for the prediction of model parameters characterizing individual subjects. As the authors of this study are working particularly on binaural aspects, a simple model of binaural hearing was exemplarily used for the present proof of concept. The chosen model can be fit to accurately simulate individual tone-in-noise detection sensitivity (Encke & Dietz, 2021).

Methods

Model-based selection framework

The method presented in Herrmann and Dietz (2021) can be separated into two parts: An analysis module with likelihood-based parameter estimation, and a stimulus-selection module running in parallel to the measurement.

The analysis module can also be used on data that was not taken using model-based steering. It only requires the model to operate as an artificial observer, meaning that it receives the same stimuli as real subjects, and provides an output that can be analyzed in the same way as the experimental data. All data is compared to pre-calculated model predictions, based on a selected set of parameter combinations. The dimensionality of this table equals the sum of stimulus- and model parameters. The comparison of data and table yields a multi-dimensional likelihood table. The compound likelihood of each model parameter is calculated by summing the likelihood values over all other parameters.

With the stimulus-selection module, the chosen stimulus is (based on the current model parameter estimations) expected to provide the most information for refining the model parameter estimates. The procedure chooses the stimulus condition that cause the largest reduction in the confidence range of likelihood over the model parameters. If one parameter is diagnostically more interesting than others, the measurement can be steered towards minimizing the confidence range of this particular parameter, or to give it a higher weighting. Here, however, the unweighted sum of all confidence intervals in units of discretization steps is minimized. To quantify the width of the confidence intervals, a function of the form

$$-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2} \tag{1}$$

was fitted to the summed log-likelihood, corresponding to a Gaussian fit of the likelihood. The parameter σ controls the width of the function. It is used as a marker of confidence in the model parameter assessment. It will be referred to here as the confidence range. More details can be found in Herrmann and Dietz (2021).

Auditory model

As noted in the Introduction, an accurate model is a crucial prerequisite for using the model-based selection framework. For this proof of concept, we opted for the binaural processing model of Encke and Dietz (2021). It can predict correct rates of tone-in-noise detection for a variety of dichotic and diotic stimuli. It consists of a monaural and a binaural branch. The monaural branch is sensitive to differences in energy between the reference and the target signal. The sensitivity is inversely proportional to model parameter σ_{mon}. The binaural branch is based on the difference between the Fisher's z-transformed complex correlation coefficients of a reference signal and a target signal As this transformation would result in infinite sensitivity to divergence of a fully coherent signal, which is not observed in the auditory system, the parameter $\hat{\rho}$ ($0<\hat{\rho}<1$) was introduced before z-transformation, thus limiting maximum sensitivity. As in the monaural branch, model parameter σ_{bin} is inversely proportional to binaural sensitivity, i.e. to the Euclidian distance between the z-transformed complex correlation coefficient of target and reference.

Predicted detection thresholds for the $N_\pi S_0$ and the $N_0 S_\pi$ condition are the same with this model. This is not the case in behavioral data (e.g. Hirsh, 1948). To account for this difference, we introduced a fourth parameter $\Delta\hat{\rho}$ into the model. It represents the linear increase in $\hat{\rho}$ when changing from anti-phasic to in-phase noise.

Model predictions are shown in Figure 1. In each panel, one model parameter was varied, while the other three parameters were set to a fixed value in the center of their respective range. As described above, each model parameter introduces changes to specific stimulus conditions, whereas others are not affected.

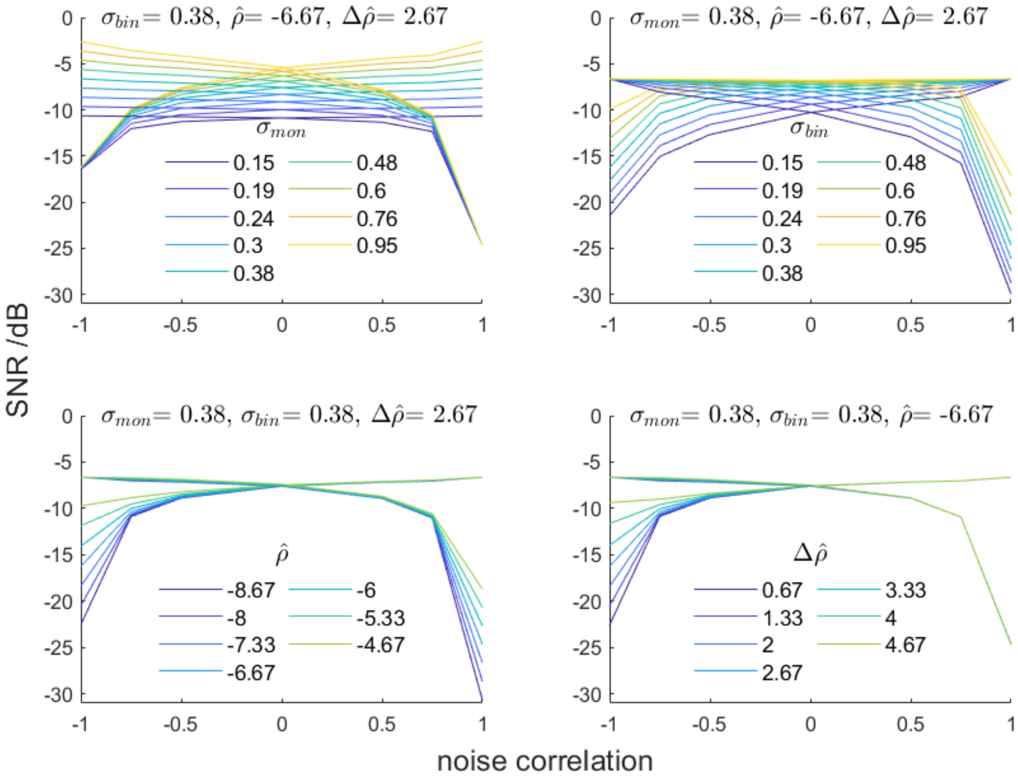

Figure 1 — Model predictions (signal-to-noise ratio corresponding to 79.4% correct) for different model instances. In each panel, one model parameter was varied (color coding), while the other three parameters were set to a fixed value in the center of their respective range.

Measurements

Five young (age: 20-26 years) participants (three female, two male) conducted the experiments with informed consent (approved by the ethics committee of the University of Oldenburg). The listeners received monetary compensation for the time spent on the experiments. Self-reported normal hearing was verified by pure tone audiometry. None of the listeners had hearing thresholds exceeding 20 dB HL and there was no more than 10 dB difference in hearing threshold between the two ears at any octave frequency between 125 Hz –and 10 kHz. A training phase to familiarize participants with the task preceded the experiments.

Tasks and stimuli

The study consisted of two parts. All subjects participated in the same tone-in-noise detection task with (1) an adaptive staircase procedure and (2) the model-based steering procedure. A four interval, two alternatives forced-choice experiment was conducted. Three intervals contained only the noise with a bandwidth of 100 Hz (spectrally rectangular band-pass white noise), centered around 250 Hz. The second or third interval additionally contained a pure tone. This pure tone of 250 Hz was either inter-aurally in phase, or differed in phase by 180 degrees. The noise's interaural correlation ρ ranged from anti-correlated to fully correlated (-1, -0.75, -0.5, 0, 0.5, 0.75, 1). The stimuli were chosen to be comparable to the those used in Robinson and Jeffress (1963). The duration of the intervals was 0.6 s, each with a pause of 0.2 s between intervals. A cosine rise-and-fall window of 20 ms was applied to the noise and pure tone separately. The tone started when the noise was at full amplitude. The sound-pressure level of the noise was fixed at 67 dB, whereas the tone level was varied adaptively during both experiments, as described below.

The listeners sat in a sound-attenuating booth on a comfortable chair in front of a computer screen and a computer keyboard. The signals were transmitted to an external audio interface (ADI-2 DAC FS, RME, Heimhausen, Germany) and presented using circumaural headphones (HD650, Sennheiser electronic GmbH, Wedemark, Germany). To visually support the temporal sequence, four rectangles lit up on the screen in succession during the four intervals. The participants' task was to decide whether the second or the third interval differed from the first and last "cueing" intervals. Responses could only be given after the fourth interval, and were then entered by pressing the number '2' or '3' on the keyboard. The button press was followed by visual feedback on the screen indicating whether the choice was correct. After a delay of 250 ms, the next trial was presented.

Adaptive staircase procedure

The first portion of the experiments was a standard adaptive staircase procedure varying the tone level following a 1-up 3-down rule converging to 79.4% correct responses (Levitt, 1971). The initial step size of 6 dB was halved every two reversals, until a step size of 1.5 dB was reached and lasted for a total of 8 reversals. Runs under the 14 stimulus conditions (seven noise correlations, each with two tone phases) were presented in random order. Each condition was presented five times to the participants. Whenever possible, a full set of 14 conditions was measured on the same day.

After completion of data collection, the analysis module with likelihood-based parameter estimation was applied to assess the most likely model parameters underlying these results. For visualization of the measured data, and for a comparison with the model predictions of the analysis module, detection thresholds corresponding to 79.4% correct responses were computed from the average of the last eight reversals of the adaptive tracks.

Model-steered procedure

In the second portion of the experiment, the measurement procedure was steered by the model-based selection approach introduced above. The range and discretization steps of the model parameters needed to be confined prior to the measurement phase. The model was run for all combinations of possible model parameters (model instances), and all combinations of possible stimulus parameters (stimulus condition). To reduce the amount of data, a logistic function was fitted to the model outcome (signal-to-noise ratio) over level for each combination of model instance and stimulus condition. These thresholds and slopes were saved in what is referred to as the model table. During the model-steering procedure, only the model outcome stored in this model table was available for the likelihood-fitting.

Depending on how the parameters influenced the model outcome, the relation between the possible values was chosen differently. For σ_{mon} and $\sigma_{bin,}$ factorial steps of $\sqrt[3]{2}$ (one third octave) ranging from 0.15 to 0.96 were chosen. Parameters $\hat{\rho}$ and $\Delta\hat{\rho}$ were chosen with steps of 2/3. Weighted by the noise correlation, they were added and are used as the exponent in

$$\left(1-2^{(\hat{\rho}+\Delta\hat{\rho})}\right)\rho \qquad (2)$$

as multiplier <1 to reduce the nominal signal correlation ρ prior to Fisher's z-transformation. Ranges from -26/3 to – 14/3 for $\hat{\rho}$ and 2/3 to 14/3 for $\Delta\hat{\rho}$ were chosen. To ensure that the expected model parameters of each subject were covered by the range of each model parameter provided, several piloting trials were necessary. The effects of changes in each of the model parameters are shown in Figure 1.

With human subjects, unlike artificial subjects switching between perceptually differing stimulus conditions between single trials leads to less reliable responses and poorer immediate performance (e.g. Taylor & Rohrer, 2010). To circumvent this, two additions were made to the original procedure. First, the measurement phase was split into several measurement blocks, each with a fixed number of repetitions of the same stimulus instance. For this study, 28 blocks of 30 repetitions of the same condition (but varying level corresponding to the point of maximal expected information) were completed by the subjects. After each block, the framework computed the next stimulus to be presented. Second, the first two repetitions of each block were carried out merely to permit familiarity with the new stimulus condition, but were neither saved nor used for the steering procedure. With this, a total of 840 trials were presented, of which 784 (28 blocks x 28 repetitions) were stored. The first four blocks were measured under pre-defined conditions before the likelihood-based measurement steering algorithm started. This was to initialize the model with a good database for the selection of the subsequent stimulus conditions. The conditions chosen for these initial blocks were: one purely monaural condition (N_0S_0), the two extreme binaural conditions (N_0S_π and $N_\pi S_0$), plus one intermediate condition ($N_{\rho=0.75}S_\pi$).

The choice of suitable initialization blocks also required knowledge acquired during the piloting of the study. With the model-based selection procedure, the accuracy of the model parameter estimation can be tracked and used to terminate the experiments. For the present 'proof-of-concept type' study, no termination criterion was set. With such a termination criterion, the measurement ends when the desired confidence range is reached in all of the model parameters. Weighting some parameters more strongly than others for the stimulus selection is also possible. To allow for comparisons between the two procedures, the 784 trials in the model-steered procedure were oriented on the number of trials in one full set of conditions in the adaptive procedure (on average ~750 trials, depending on measurement set and subject).

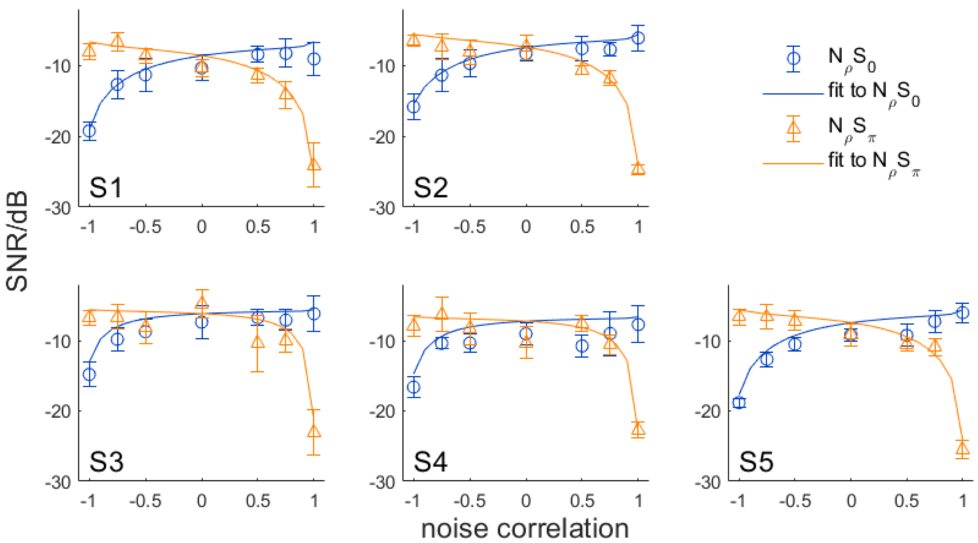

Figure 2 — Tone-in-noise detection thresholds of the five subjects. The orange triangles represent thresholds for stimuli with anti-phasic tones, the blue circles for tones that were inter-aurally in phase. The standard deviations across the five trials of each condition are represented as error bars. The solid lines represent the model predictions for the fit with the model parameters presented in Table I.

Results

Adaptive staircase procedure

The tone-in-noise detection thresholds are shown in Figure 2. Using the likelihood-based analysis, model parameters corresponding best to the subjects' data were obtained. The resulting model predictions are displayed as solid lines in the same figure, and show the estimated signal-to-noise ratio for 79.4% correct. Model parameters obtained by the analysis module can be found in Table I.

As expected, the thresholds for the conditions without binaural cues (N_0S_0, the right-most blue data point and $N_\pi S_\pi$, the left-most orange data point) are the highest. Thresholds improved with increasing average IPD difference between masker and target, until the lowest thresholds were obtained for $N_\pi S_0$ (leftmost blue data point) and N_0S_π (rightmost orange data point). Within the latter condition, all subjects reached the lowest of their thresholds.

The model fit captures the behavior of all subjects, with only small deviations for single stimulus conditions. The coefficient of determination R^2 ranged between 79.9 % for S4 and 96.5 % for S2 and was averaged 89.8 %. Lower values of σ_{mon} correspond to lower thresholds in the monaural conditions. Lower values of σ_{bin} correspond to lower thresholds in those conditions with binaural cues. As described in the Methods section, mainly the thresholds for $N_\pi S_0$ and N_0S_π are affected by parameter $\hat{\rho}$, whereas $\Delta\hat{\rho}$ enables fitting the difference between $N_\pi S_0$ and N_0S_π.

Table I — Model parameters estimated by the analysis module for data of the adaptive staircase procedure and the model-steered procedure. The deviation between the two procedures is color-coded with light grey for one bin, middle grey: two bins, and dark grey corresponding to three bins.

subject	procedure	model parameter			
		σ_{mon}	σ_{bin}	$\hat{\rho}$	$\Delta\hat{\rho}$
S1	adaptive	0.38	0.24	-6.00	2.00
	model-steered	0.38	0.30	-6.00	2.00
S2	adaptive	0.48	0.30	-6.00	2.67
	model-steered	0.48	0.60	-7.33	2.67
S3	adaptive	0.48	0.60	-6.67	2.67
	model-steered	0.48	0.48	-6.67	2.67
S4	adaptive	0.38	0.48	-6.67	2.67
	model-steered	0.60	0.48	-6.67	1.33
S5	adaptive	0.48	0.30	-6.00	2.00
	model-steered	0.48	0.48	-6.67	2.00

Model-steered procedure

The model table was pre-calculated overnight on a regular i5 laptop. The table contains the thresholds and slopes of the logistic functions representing the model output for the combination of the 2x7 stimulus conditions and the 9x9x7x7 model instances, leading to a total of 55566 model calls for each stimulus level.

When using the stimulus-selection module, model parameters were estimated for every trial. The development of the normalized likelihood of the four parameters over trials and the stimuli chosen by the procedure are shown in examples for subject S4 in Figure 3. In each of the upper four panels, the likelihood values are summed over the other three parameters and independently normalized by the mean of the summed likelihood for each model parameter and each trial. Over the course of the trials, the likelihood distribution reduced in width around the maximal value.

The first stimulus condition in the experiment (N_0S_π) did not deliver information on the monaural threshold. For this reason, the estimation of σ_{mon} only starts refining with the second block (N_0S_0). Similarly, parameter $\Delta\hat{\rho}$ (the difference between N_0S_π and $N_\pi S_0$) can only be estimated starting with the first trials of $N_\pi S_0$ in block number three. To ensure a good start to the measurement procedure without too much emphasis on one single stimulus condition, the first four blocks were set as initialization blocks. Starting with trial number 113 (at the dashed black line), the stimulus-selection module selected different stimulus instances, with an emphasis on N_0S_π and $N_{\rho=0.75}S_\pi$, and, to a lesser degree, on $N_\pi S_0$ and $N_\pi S_\pi$. Noise correlation values between those were almost not chosen at all. Comparable patterns were also found for the other subjects. The final estimates of the model parameters with the data acquired using the model-steered procedure are shown in Table 1. The respective model parameter estimates for the data of the two measurement procedures are very similar. One bin deviation (light grey) is not significant, as it can result from discretization following an arbitrarily tiny difference in the data. Also, confidence intervals are of the order of one bin (0.69 bins, see next subsection). Darker grey shadings indicate stronger differences between the estimates of the two procedures. In those cells without shading, the parameter estimations did not differ. Only for subject S2 did the parameter estimations differ by more than two steps on the grid (bins) of possible values. Across subjects, estimates differed by up to three bins for σ_{bin}, but only by up to two bins for the other three parameters. Differences across subjects are only marginally larger than within-subject differences obtained with the two methods.

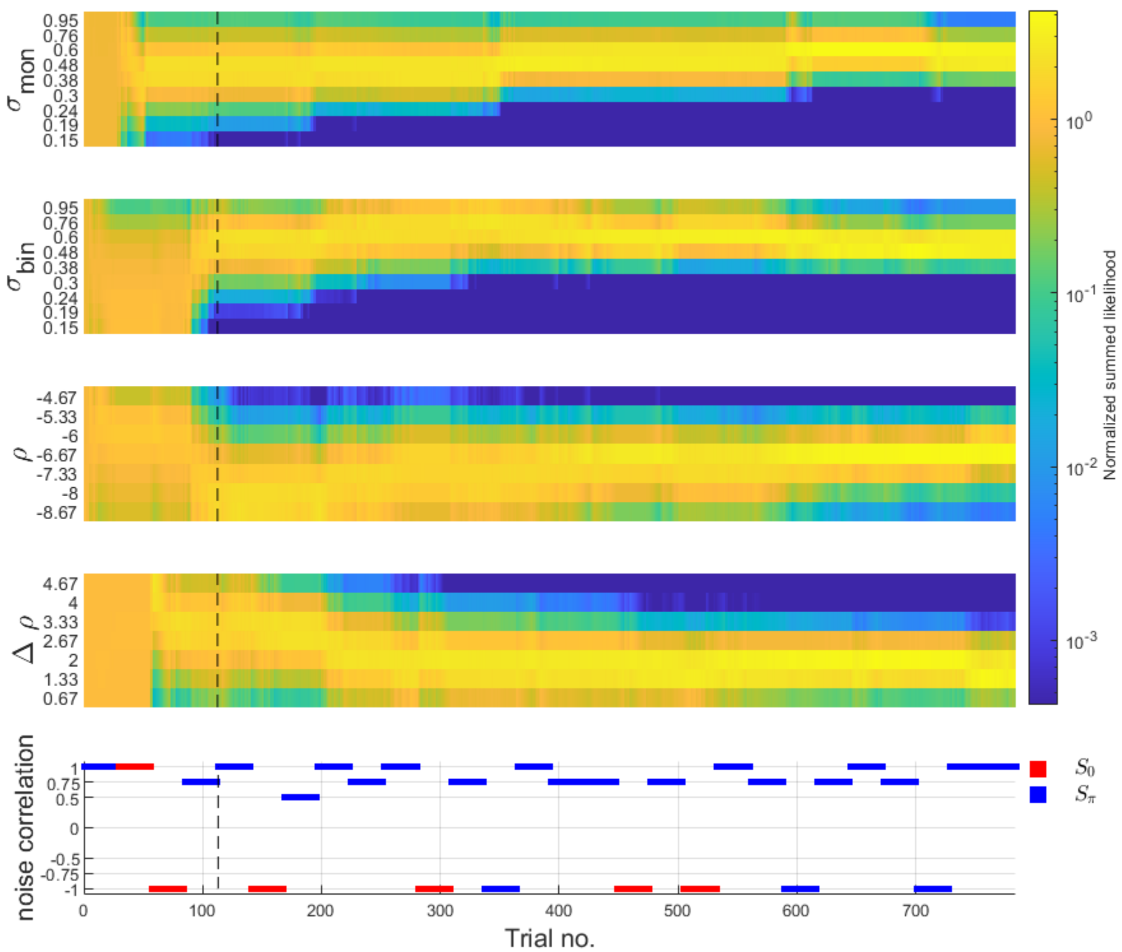

Figure 3 — Mean-normalized likelihood (sum over the other three parameters) of the four parameters across trials in the upper four panels, and the stimuli chosen by the procedure across trials in the bottom panel for subject S4. The dashed black line indicates the start of the model steering with trial number 113.

The confidence ranges can be estimated from Figure 3. Even though the confidence ranges did not become smaller with every single trial, overall, the accuracy of the estimations increased. For instance, the mean confidence range over the four parameters for subject S4 decreased from 4.28 bins at the start of the model-steering to 0.72 bins after the last trial. Comparable decreases were also found for the other subjects. After the final trial, the procedure reached a mean accuracy between 0.64 bins and 0.72 bins for the different subjects (mean: 0.69 bins).

Comparison of the two procedures

Figure 4 shows the mean confidence ranges across the four model parameters over trials for the two data sets, the adaptive and the model-steered procedures. With the latter, only 874 trials were recorded. The confidence ranges are only shown after each full set of 14 conditions for the adaptive procedure. These sets differed slightly in the number of trials for the five subjects. In general, a decrease of confidence ranges over trials was observed, with a steeper decrease in the model-steered data. For instance, for subject S1, the mean confidence range was 1.63 bins after the first set of the adaptive procedure (693 trials). The value of 1.63 bins or smaller was initially reached after 302 trials with the model-steered procedure. The ratio of 2.29 indicates that the same confidence range was achieved more than twice as fast with the model-based steering method. To reach the same confidence range using the adaptive procedure, 1.9 to 3.7 times more trials were necessary than with the model-steering procedure.

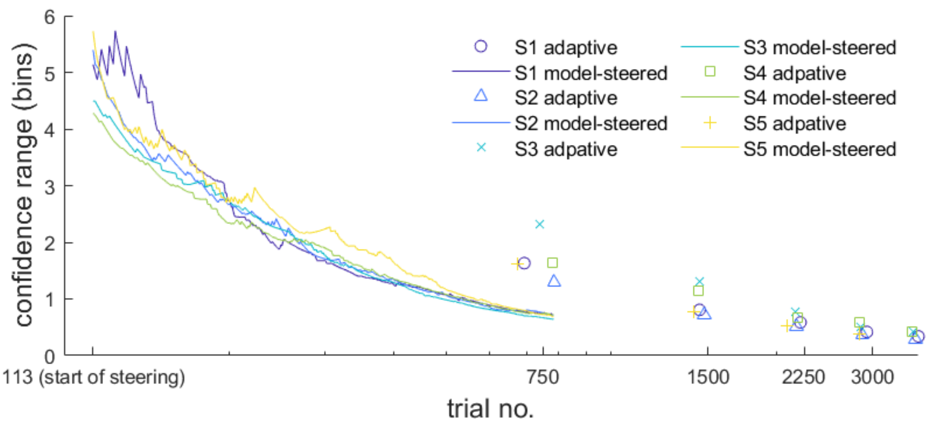

Figure 4 — Mean of confidence ranges (in bins) averaged across the four model parameters over trials for subjects S1 to S5. The estimates for the model-steered procedure are depicted with lines, for the adaptive procedure with symbols. Color and marker shape vary for the individual subjects.

Discussion

This study sought to affirm the feasibility of model-based experiment steering after a preceding study by Herrmann and Dietz (2021) had concluded that there would be a theoretical advantage of the proposed procedure over sequential measure-and-fit approaches. In the current study, the model-steered procedure was tested on real, but non-impaired subjects, instead of an artificial patient simulated on a computer. This attempt was successful for two reasons. First, the estimated model parameters were sufficiently close to those obtained from the results of the standard adaptive procedure (on average deviating by 0.70 bins, corresponding to an average threshold differences of 0.98 dB). Second, the same accuracy in model parameter estimates was obtained 1.7 to 3.7 times faster than with the standard method.

The proposed measurement procedure can assist in linking data to the underlying pathology, or to a parametric description of the individuals' hearing abilities. The procedure will steer towards those measurements that can disentangle different causes of the observed behavior, even in the complex auditory processing chain. As a prerequisite for this becoming reality in clinical settings, models with high diagnostic resolution need to be developed. In the current study, an existing simple model of binaural processing was used, but slightly adapted as a first attempt to characterize a subject in the most time-efficient way. Even though the diagnostic value of the model parameters is not clear, it served as proof of concept.

The duration of measurements is limited in clinical settings. Keeping measurement times as short as possible is, however, also of importance for another reason: With longer measurement times, unaccounted factors influence the data. Fatigue, attention, motivation or effects specific to single measurement days may potentially confound the parameter evaluations. Using the model-steering procedure, confidence ranges of less than 0.69 bins, corresponding on average to threshold uncertainties of approximately 1 dB, were reached for the four model parameters after less than 1.5 hours.

Choosing meaningful ranges and discretization for the model parameters remains a critical point. Extensive piloting with adjustments to the ranges and step sizes of the four parameters preceded data collection. Each step must influence the model outcome for at least one stimulus instance. In the best case, each step leads to similarly large changes in model predictions. In Figure 1, it can be seen that changing σ_{mon} by one step always leads to changes in the estimated signal-to-noise ratio of about 1 dB. Similar changes are observed for σ_{bin} but in other stimulus conditions. Increasing or decreasing parameters $\hat{\rho}$ and $\Delta\hat{\rho}$ by one step always leads to a change of about 2 dB, but influences fewer stimulus conditions. When comparing the selected stimulus conditions (see the bottom panel in Figure 3) to the changes in the model prediction in Figure 1, it becomes obvious which stimulus conditions provide the most information about each of the model parameters. $N_{\pi}S_{\pi}$ is chosen, as it only depends on σ_{mon}. The frequently chosen condition $N_{\rho=0.75}S_{\pi}$, for example, mainly informs about $\sigma_{bin.}$ However, the more complex the models are, the more difficult it is to comprehend these relationships.

Matching the effect size of parameter steps is also important in the light of co-dependencies between model parameters. Preferably, changes by one discrete step in one parameter should not force another co-dependent parameter to change by more than one step, otherwise the undesirable effect may occur producing several maxima in the likelihood over the values of the co-dependent model parameter. For example, if parameter X changes by one step and causes a counter-reaction of parameter Y by two steps, the likelihood function of parameter Y may not follow a Gaussian bell shape. High likelihood values that each correspond to a particular value of X might alternate with low likelihood values that have no direct correspondence to a particular value of X. Arguably, fitting a Gaussian to Y or even trying to reduce its confidence interval is problematic in such a case. It has also proven to be important that estimated parameters do not reach the boundary of the parameter range of the previously stored model table. In this case, the fit with an inverted parabola cannot represent the log-likelihood values over the parameters very well. To fit the data best, the apex of the parabola would possibly be outside the boundaries. The steepness (given by parameter a) would be very small, resulting in confidence ranges spanning the entire possible range of parameters. Such corrupted confidence ranges lead to the choice of non-optimal next stimulus conditions. An additional advantage of matching the effect size of parameter steps is that it allows the steering procedure to minimize the unweighted sum of confidence ranges, as measured in numbers of steps or bins. The procedure is then expected to provide similar accuracy for all parameters. With ill-matched parameters, the model steering might be biased towards minimizing the confidence ranges of some model parameters more than others.

The computational demand of the procedure can be a limiting factor. Model tables cannot be chosen with arbitrarily high resolution, as between the measurement blocks, the model-steering module needs to load the whole table and to choose the next stimulus by computing the estimated confidence ranges for each stimulus instance. With too-extensive model tables, the computational demands become too high. Two improvements helped to reduce this. First, substituting the numerical model with an analytical model reduced the computation time of each model call. Second, while the model used by Herrmann and Dietz (2021) provided the expected target interval number, just like for a real subject, the output of the new model is directly the d’ value that can be converted to the correct rate. While this is not the response of a human subject, it is more helpful because it does not require hundreds of repetitions for the same condition just to obtain a reliable correct rate estimate.

One of the main concerns remains the choice of an accurate model with diagnostic value. The approach with an auditory processing model requires the faithful simulation of the whole chain from stimulus presentation, through internal processing to the subject’s response, or to other measurement data. We were able to perform a proof-of-concept, but could only characterize those aspects that are relevant for tone-in-noise detection sensitivity at one frequency and only for normal-hearing subjects. Some of the four free parameters are expected to characterize the consequences of hearing impairment (e.g. Bernstein & Trahiotis, 2018). Other model parameters such as filter bandwidth are fixed, however, and this cannot serve as a realistic model for patients with outer-hair-cell impairments. Of course, filter bandwidth can be an additional parameter to fit, as already demonstrated (Herrmann & Dietz, 2021), and most other specific extensions are also expected to be compatible with the approach. The problem is the number of parameters that quickly arise (e.g. Verhulst et al., 2018), especially as many of the parameters may differ from frequency to frequency. Even some frequency-independent parameters, e.g., the endocochlear potential, will influence performance differently across frequency, making it non-trivial to fit one parameter based on prerecorded individual data (Panda et al., 2014). Abstract models that even avoid a simulation of auditory processing may be more realistic candidates for model-steered profiling if, instead of a detailed diagnosis, the focus of interest is rather on the consequences. Ideally, each model parameter should directly relate to a practical consequence, e.g., it can be a hearing-aid fitting parameter (similar to Plomp, 1978).

To summarize, the distant goal of diagnosing the causes of a person’s hearing impairment in a more objective and more time-efficient way has not yet been reached. First steps in this direction were, however, made with the model-based approach presented in this study. Characterization of individuals in terms of abstract parameters that influence hearing-aid fitting or the choice of hearing implants is possible. Scientifically, both the likelihood-based fitting and the model-based steering foster a deeper understanding of the inner mechanics of the models used. The procedure also offers insights into its interaction with fitting tools, measurement procedures, and subject peculiarities that are not captured by the model. Specifically, as argued by Herrmann and Dietz (2021), tracing why the model chooses certain stimuli and in which order, is very informative, even for an improvement of conventional manual measurement selection. It also makes it possible to fully understand the impact of each model parameter in general, and of each parameter’s discretization steps. The procedure thus provides new perspectives for the design of diagnostic models and experiments.

Acknowledgements

This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme grant agreement no. 716800 (ERC Starting Grant to Mathias Dietz).

References

Bernstein, L. R., & Trahiotis, C. (2018). Effects of interaural delay, center frequency, and no more than "slight" hearing loss on precision of binaural processing: Empirical data and quantitative modeling. *The Journal of the Acoustical Society of America*, *144*(1), 292. https://doi.org/10.1121/1.5046515

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, *111*(6), 2801–2810. https://doi.org/10.1121/1.1479152

Encke, J., & Dietz, M. (2021, November 8). *The complex-valued correlation coefficient accounts for binaural detection*. http://arxiv.org/pdf/2111.04637v2

Herrmann, S., & Dietz, M. (2021). Model-based selection of most informative diagnostic tests and test parameters. *Acta Acustica*, *5*, 51. https://doi.org/10.1051/aacus/2021043

Hirsh, I. J. (1948). The Influence of Interaural Phase on Interaural Summation and Inhibition. *The Journal of the Acoustical Society of America*, *20*(4), 536–544. https://doi.org/10.1121/1.1906407

Hoth, S., & Baljic, I. (2017). Current audiological diagnostics. *GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery*(16). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5738938/pdf/CTO-16-09.pdf

Klug, J., Schmors, L., Ashida, G., & Dietz, M. (2020). Neural rate difference model can account for lateralization of high-frequency stimuli. *The Journal of the Acoustical Society of America*, *148*(2), 678. https://doi.org/10.1121/10.0001602

Levitt, H. (1971). Transformed Up-down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*(49), 467–477. https://doi.org/10.1121/1.1912375

Meddis, R. (2006). Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *The Journal of the Acoustical Society of America*, *119*(1), 406–417. https://doi.org/10.1121/1.2139628

Panda, M. R., Lecluyse, W., Tan, C. M., Jürgens, T., & Meddis, R. (2014). Hearing dummies: Individualized computer models of hearing impairment. *International Journal of Audiology*, *53*(10), 699–709. https://doi.org/10.3109/14992027.2014.917206

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, *63*(2), 533–549. https://doi.org/10.1121/1.381753

Robinson, D. E., & Jeffress, L. A. (1963). Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals. *The Journal of the Acoustical Society of America*, *35*(12), 1947–1952. https://doi.org/10.1121/1.1918864

Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., & Dau, T. (2018). Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss. *Trends in Hearing*, *22*, 2331216518807400. https://doi.org/10.1177/2331216518807400

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837–848. https://doi.org/10.1002/acp.1598

Verhulst, S., Altoè, A., & Vasilkov, V. (2018). Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing Research*, *360*, 55–75. https://doi.org/10.1016/j.heares.2017.12.018