(54) Title: METHOD OF OPTIMIZING PARAMETERS IN A HEARING AID SYSTEM AND A HEARING AID SYSTEM

(57) Abstract: A method of optimizing parameters in a hearing aid system (100) with respect to a user's preference and a hearing aid system (100) adapted to carry out such a method.
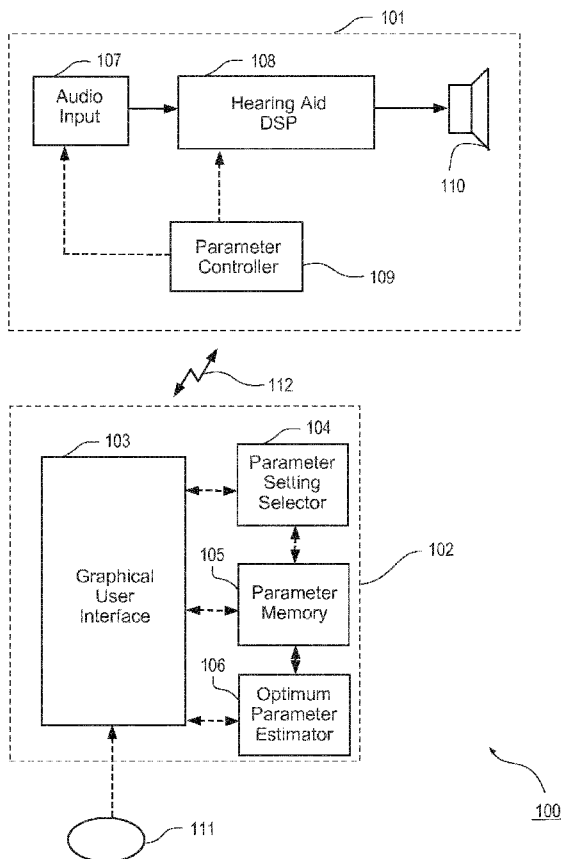
Fig. 1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHOD OF OPTIMIZING PARAMETERS IN A HEARING AID SYSTEM AND A HEARING AID SYSTEM

The present invention relates to a method of optimizing parameters in a hearing aid system. The invention also relates to a hearing aid system adapted for optimizing parameters.

BACKGROUND OF THE INVENTION

Within the context of the present disclosure a hearing aid can be understood as a small, battery-powered, microelectronic device designed to be worn behind or in the human ear by a hearing-impaired user. Prior to use, the hearing aid is adjusted by a hearing aid fitter according to a prescription. The prescription is based on a hearing test, resulting in a so-called audiogram, of the performance of the hearing-impaired user's unaided hearing. The prescription is developed to reach a setting where the hearing aid will alleviate a hearing loss by amplifying sound at frequencies in those parts of the audible frequency range where the user suffers a hearing deficit. A hearing aid comprises one or more microphones, a battery, a microelectronic circuit comprising a signal processor adapted to provide amplification in those parts of the audible frequency range where the user suffers a hearing deficit, and an acoustic output transducer. The signal processor is preferably a digital signal processor. The hearing aid is enclosed in a casing suitable for fitting behind or in a human ear.

Within the present context a hearing aid system may comprise a single hearing aid (a so called monaural hearing aid system) or comprise two hearing aids, one for each ear of the hearing aid user (a so called binaural hearing aid system). Furthermore the hearing aid system may comprise an external device, such as a smart phone having software applications adapted to interact with other devices of the hearing aid system. Thus within the present context the term "hearing aid system device" may denote a hearing aid or an external device.

Generally a hearing aid system according to the invention is understood as meaning any system which provides an output signal that can be perceived as an acoustic signal by a user or contributes to providing such an output signal and which has means which are used to compensate for an individual hearing loss of the user or contribute to

compensating for the hearing loss of the user. These systems may comprise hearing aids which can be worn on the body or on the head, in particular on or in the ear, and can be fully or partially implanted. However, some devices whose main aim is not to compensate for a hearing loss may nevertheless be considered a hearing aid system, for example consumer electronic devices (televisions, hi-fi systems, mobile phones, MP3 players etc.) provided they have measures for compensating for an individual hearing loss.

It is well known within the art of hearing aid systems that most users will benefit from a hearing aid programming (this process may also be denoted fitting) that takes the user's personal preferences into account. This type of fine tuning or optimization of the hearing aid system settings may also be denoted customization. It is however also well known that the process of customization is a very challenging one.

One problem with customization is that it may be very difficult for a user to explain in words what types of signal processing and the resulting sounds that are preferred.

Customization may generally be advantageous with respect to basically all the various types of signal processing that are carried out in a hearing aid system. Thus customization may be relevant for e.g. noise reduction as well as for classification of the sound environment.

EP-B1-1946609 discloses a method for optimization of hearing aid parameters. The method is based on Bayesian incremental preference elicitation whereby at least one signal processing parameter is adjusted in response to a user adjustment. According to a specific embodiment the user adjustment is simply an indication of user dissent.

EP-B1-1946609 is complicated in so far that it applies a parameterized approach in order to model the user's unknown internal response function (i.e. the user's preference), because it is very difficult to find a suitable parameterized model that suits the great variety of hearing aid system users unknown internal response functions.

Furthermore EP-B1-1946609 is complicated because the processing and memory requirements are very high, especially for hearing aid systems that generally have limited processing and memory resources.

It is therefore a feature of the present invention to provide an improved method of optimizing a hearing aid system setting with respect to both user satisfaction and requirements to processing and memory resources.

It is another feature of the present invention to provide a hearing aid system with
5   improved means for optimizing a hearing aid system setting.

Additionally the inventor has found that internally generated sounds that are used for providing comfort, be it for masking undesired sounds or just for causing a relaxing experience, may benefit significantly from customization.

In the context of the present disclosure, a relaxing sound should be understood as a
10   sound having a quality whereby it is easy to relax and be relieved of e.g. stress and anxiety when subjected to it. Traditional music is one example of relaxing sound while noise is most often used to refer to a sound that is not relaxing.

In the context of the present disclosure, a relaxing sound may especially be understood as a sound adapted for relieving tinnitus.

15   US-B2-6816599 discloses one type of relaxing sound, that can be generated by a music synthesizer in a way that is very well suited for implementation in e.g. a hearing aid.

US-6047074 discloses a hearing aid that can also be utilized for tinnitus therapy, wherein a useful digital signal, derived from the output signal from the hearing aid input transducer, can be evaluated in terms of its intensity, its spectral distribution
20   and/or its time structure such that an oppositely directed (compensating) behavior can be achieved. Hereby the signals for tinnitus therapy can be activated only when no useful signal is present. Arbitrary transition times between end of the useful signal and beginning of the signals for tinnitus therapy can thereby be set. When a longer quiet pause occurs, then the masking signal is slowly mixed in and thus drowns out the
25   disturbing tinnitus noise. It is also disclosed that melodic sound sequences or other tones can be used to mask the tinnitus

One problem with generating relaxing sounds is that it may be very difficult for a user to explain in words what type of sounds are perceived as relaxing.

This is especially critical if the user desires to use the relaxing sounds in order to draw
30   his attention away from e.g. a perceived tinnitus tone.

It is therefore a feature of the present invention to provide an improved method for customizing the generation of relaxing sound.

It is another feature of the present invention to provide a hearing aid system with improved means for customizing the generation of relaxing sound.

5   SUMMARY OF THE INVENTION

The invention, in a first aspect, provides a method according to claim 1.

This provides an improved method for optimizing or customizing parameters in a hearing aid system.

The invention, in a second aspect, provides a method according to claim 18.

10   This provides an improved method for optimizing or customizing parameters in a hearing aid system.

The invention, in a third aspect, provides a method according to claim 19.

This provides an improved method for optimizing or customizing parameters in a hearing aid system.

15   The invention, in a third aspect, provides a hearing aid system according to claim 20.

This provides an improved hearing aid system adapted to customize or optimize parameters in a hearing aid system.

Further advantageous features appear from the dependent claims.

Still other features of the present invention will become apparent to those skilled in the

20   art from the following description wherein the invention will be explained in greater detail.

BRIEF DESCRIPTION OF THE DRAWINGS

By way of example, there is shown and described a preferred embodiment of this invention. As will be realized, the invention is capable of other embodiments, and its

25   several details are capable of modification in various, obvious aspects all without departing from the invention. Accordingly, the drawings and descriptions will be regarded as illustrative in nature and not as restrictive. In the drawings:

Fig. 1   illustrates highly schematically a hearing aid system according to an
embodiment of the invention.

DETAILED DESCRIPTION

In the present context the term relaxing sound represents sound that is generated
synthetically in a hearing aid in order to help a user to concentrate, to feel more relaxed
and comfortable, to reduce stress and to feel less anxious.

In one aspect the relaxing sound can achieve this by masking or by drawing a user's
attention away from unwanted and disturbing sounds. In another aspect it has been
found that the relaxing sound can by itself help to achieve this independent on whether
unwanted and disturbing sounds are present or not.

Generally the inventor has found that it provides a significant improvement for the user
if a relaxing sound can be adapted to the user's current preferences (i.e. customized)
because the inventor has found that the user's preferences may change significantly up
to several times during a day, thus the preferences may depend on the time of day
(morning, afternoon or evening) or the user's mood or the type of activity the user is
engaged in.

As a consequence of the varying preferences of many users the inventor has also found
that it provides a significant improvement for the user if the customization can be
carried out without having to spend too much time listening to various types of relaxing
sound.

As an additional consequence of the varying preferences of many users the inventor has
found that it provides a significant improvement for the user if the customization can be
carried out using only the hearing aid system with its limited processing resources,
because this allows the customization to be carried out anywhere and at any time.

Furthermore the inventor has found that it is of significant importance that the
customization can be carried out without requiring the user to interact with the hearing
aid system in a complex manner.

Especially, the inventor has found that analytical expressions allowing hearing aid
system customization to be carried out with beforehand unseen processing efficiency

can be derived if a hearing aid system user is prompted to compare two hearing aid system settings and rate how much one of the settings is preferred above the other.

As described above the present invention is especially advantageous for customization of relaxing sounds.

5   One example of artificially generated relaxing sounds that are suitable for implementation in a hearing aid system can be found e.g. in WO-A1-02/41296

Further the inventor has found that artificially generated relaxing sounds such as those disclosed in WO-A1-02/41296 are advantageously customized by optimizing the ranges wherein parameters used to control the generation of sound are allowed to vary

10   pseudo-randomly. However, as one alternative to that customization approach the inventor has found that optimization of the harmonics characteristics of the generated relaxing sounds may provide a significant improvement for the user.

Reference is first made to a method of customizing relaxing sound according to a first embodiment of the invention.

15   In a first step a set of parameters are selected from the group of parameters that controls generation of relaxing sound by the hearing aid system.

The parameters are selected such that they, when varied over their allowed range, are able to provide a multitude of relaxing sounds that are perceived as having a significant variation.

20   According to one variation of the present embodiment and having reference to WO-A1-02/41296 the set of parameters comprises the specific harmonics added to the signals generated by a multitude of sound generators according to the embodiments of WO-A1-02/41296. Hereby, the sound generators are customized to provide sound with the harmonic characteristics preferred by the user.

25   According to another variation of the present embodiment and again having reference to WO-A1-02/41296 the first set of parameters comprises, for one or more sound generators, the allowed range within which the frequency of the sound generator and the fade-out time may vary as controlled by the provided random numbers. Thus the inventor has found that a significant improvement for the user may be provided by

customizing the range within which a parameter value is allowed to vary, even though this variation may be randomly controlled.

In a second step a first and second set of parameter values are selected (this may also be denoted the first and second parameter value settings, or just the first and second settings), whereby a first relaxing sound is generated and provided to the user based on the first set of parameter values and a second relaxing sound is generated and provided to the user based on the second set of parameter values. According to the present embodiment the first and second sets of parameter values are selected randomly.

In variations the first and second set of parameter values need not be selected randomly. Instead the first set of parameter values is the set that was active when powering off the hearing aid system. In further variations the first and second parameter value settings may be identical.

Consider now a d-dimensional vector x containing the values of the d selected parameters that controls the generation of relaxing sound by the hearing aid system:

$$x = [x_1, \dots, x_d]^T$$

In the following a d-dimensional vector x with specific values of the d parameters may also be denoted a setting or a parameter value setting.

In a third step the user is prompted to compare the first and second relaxing sound and provide a first user response (that in the following may also be denoted an observation) that allows a determination of which of the two relaxing sounds the user prefers.

According to the present embodiment the observations comprise a graduated response whereby the user rates how much the first relaxing sound (and hereby the first parameter value setting) is preferred above the second relaxing sound (and hereby the second parameter value setting) by selecting a number from within a bounded range (that may also be denoted an interval) between zero and one, such that a user response of one implies that e.g. the first parameter value setting is indefinitely better than the second parameter value setting and zero implying that the second set is indefinitely better than the first set and that a value of one half implies that the two options are rated to be equally good.

In variations of the present embodiment the bounded range may cover basically any range such as e.g. the range from -1 to +1 or the range from -10 to +10. However, it is emphasized that generally and especially within the present context a test where a user selects either one or another setting can't be considered to provide a user's rating of the settings relative to each other. Thus within the present context when the user response takes the form of a number from a bounded range, then this range (or interval) has more than two elements.

In a fourth step the user provides a multitude of additional user responses based on a multitude of parameter value settings, wherein a set $\mathcal{X}$ of the n tested parameter value settings may be given as:

$$\mathcal{X} = \{x_i \in \mathbb{R}^d : i = 1, \dots, n\}$$

and wherein a vector of m bounded observations are related to pairs of relaxing sounds based on the settings $x_{u_k}, x_{v_k} \in \mathcal{X}$, implying that $u_k, v_k \in \{1, \dots, n\}$, such that:

$$y = [y_1, \dots, y_m]^T$$

Wherein the bounded observations are of the form

$$y_k \in \left]a; b\right[$$

The set $\mathcal{X}^*$ of all n^* possible parameter value settings is expressed as:

$$\mathcal{X}^* = \{x_r^* \in \mathbb{R}^d : r = 1, \dots, n^*\}$$

The user's unknown internal response function is denoted f and is assumed to code the user's perception of a particular sound given the setting x of the parameter values.

$$f : \mathbb{R}^d \rightarrow \mathbb{R}, x \mapsto f(x)$$

In the following a (stochastic) vector $\mathbf{f}$ is defined as containing the function values $f(x_i)$ for each of the n settings in $\mathcal{X}$ of the user's internal response function f:

$$f = [f(x_1), \dots, f(x_n)]^T$$

In a fifth step the vector of bounded observations $\mathbf{y}$ are warped to a vector of unbounded observations $\mathbf{z}$. This is done using a warping function defined by mapping a bounded, or partly bounded observation, y, to an unbounded observation, z = g(y), with:

$$g : \left]a; b\right[\rightarrow \mathbb{R}, y \mapsto g(y)$$

Suitable functions for carrying out the mapping may be selected from a group of monotonically increasing functions comprising: inverse cumulative distribution function of the Gaussian distribution, inverse sigmoid function and inverse hyperbolic tangent function.

By mapping or warping the observations (in the following these terms may be used interchangeably) in this way the performance of the customization method may be improved both with respect to the speed of convergence (i.e. the number of user responses required to find the parameter value settings that the user prefers) and with respect to robustness (i.e. the chance that the customization method is capable of providing a prediction that reflects the users internal response function). However, it is not a prerequisite for the methods of the present invention that warping is applied.

In a sixth step it is assumed that the warped observations z are given by:

$$z = f(x_u) - f(x_v) + \epsilon$$

wherein ϵ is Gaussian noise ($\epsilon \sim N(0,\sigma^2)$) that is independent and identically distributed and represents the uncertainty of the user when carrying out the graduated responses.

Based on this assumption the likelihood function for observing z can be determined directly as:

$$p(z_k | f(x_u), f(x_v), \sigma^2) = \mathcal{N}(z_k | f(x_u) - f(x_v), \sigma^2)$$

wherein z_k represents a specific warped user response, wherein σ^2 represents the variance of the user response and wherein $\mathcal{N}(z_k | f(x_u) - f(x_v), \sigma^2)$ is the single variate Gaussian distribution over the variable z_k with mean value $f(x_u) - f(x_v)$ and variance σ^2. In the following the variance σ^2 may also be denoted the likelihood hyper parameter θ_{lik}.

In a seventh step the likelihood is determined using the following new and inventive expression as:

$$p(z|f, \theta_{lik}) = \prod_{k=1}^{m} \mathcal{N}\left(z_k | f(x_{u_k}) - f(x_{v_k}), \sigma^2\right) = \mathcal{N}(z|Mf, \sigma^2 I_{m\times m}),$$

wherein $I_{m \times m}$ is an $m \times m$ identity matrix, $z = [z_1, \ldots, z_m]^T = [g(y_1), \ldots, g(y_m)]^T$, the matrix $\mathbf{M}$ is a m×n matrix, comprising only zeros except for the elements $[M]_{k,u_k} = 1$ and $[M]_{k,v_k} = -1$.

Thus $\mathcal{N}(z|Mf, \sigma^2 I_{m \times m})$ represents a multivariate Gaussian distribution over the set of user responses z with mean vector Mf and covariance matrix $\sigma^2 I_{m \times m}$.

In an eighth step a prior distribution $p(f|\mathcal{X}, \boldsymbol{\theta}_{cov})$ over the function values of the user's unknown internal response function is obtained from a zero-mean Gaussian process. This is obtained based on the fact that a zero-mean Gaussian process defines a joint distribution over a finite set of function values, f, as a multivariate Gaussian distribution

$$p(f|\mathcal{X}, \boldsymbol{\theta}_{cov}) = \mathcal{N}(f|0, K)$$

where the covariance between any *two* function values is given by a positive semi-definite covariance function, $k(x_i, x_j, \boldsymbol{\theta}_{cov})$, such that the covariance matrix $\mathbf{K}$ is determined by:

$$K = \begin{bmatrix} k(x_1, x_1, \boldsymbol{\theta}_{cov}) & \cdots & k(x_1, x_n, \boldsymbol{\theta}_{cov}) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1, \boldsymbol{\theta}_{cov}) & \cdots & k(x_n, x_n, \boldsymbol{\theta}_{cov}) \end{bmatrix}$$

and wherein k is the squared exponential covariance function (that may also be denoted the Gaussian kernel) that is defined as:

$$k(x_i, x_j) = \sigma_f Exp(-\tfrac{1}{2}(x_i - x_j)^\top L^{-1}(x_i - x_j))$$

wherein σ_f and the positive semi-definite matrix $\mathbf{L}$ are the covariance hyper parameters, that primarily determines the smoothness of the user's internal response function. The covariance hyper parameters σ_f and $\mathbf{L}$ may together be denoted $\boldsymbol{\theta}_{cov}$.

Generally the prior obtained from a Gaussian Process (that in the following may be abbreviated GP) captures the assumption about the smoothness of the users internal response function f as a function of the parameter value settings x and interplay between the different hearing aid parameters.

It is a fundamental property of a Gaussian process that you can always get the joint distribution of any finite set of function values from a function f(x) modeled by a

11

Gaussian Process. Hence, we can concatenate two vectors containing function values
and get the corresponding joint distribution, e.g. as.

$$p\left(\begin{bmatrix} f \\ f^* \end{bmatrix} \middle| \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}_{cov}\right) = \mathcal{N}\left(\begin{bmatrix} f \\ f^* \end{bmatrix} \middle| \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}\right),$$

where:

$$[K_*]_{i,s} = k(x_i, x_s^*, \boldsymbol{\theta}_{cov})$$

$$[K_{**}]_{r,s} = k(x_r^*, x_s^*, \boldsymbol{\theta}_{cov})$$

5   Wherefrom it is trivial to obtain the conditional distribution of one vector of function
values given the other vector. E.g.:

$$p(f^*|f, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}_{cov}) = \mathcal{N}(f^* | K_*^T K^{-1} f, K_{**} - K_*^T K^{-1} K_*)$$

However, in variations of the present embodiment other positive semi-definite
covariance functions may be applied instead of the Gaussian kernel. Examples of such
10   covariance functions are the γ-exponential covariance function, the Matern class of
covariance functions, the rational quadratic covariance function, periodic covariance
functions and linear covariance functions.

In a ninth step the hyper parameters $\theta = \{\theta_{cov}, \theta_{lik}\}$ are selected based on a
maximization of the expression for the marginal likelihood, whereby the likelihood of
15   the observed data is also maximized. The marginal likelihood is determined based on
the likelihood $p(z|f, \boldsymbol{\theta}_{lik})$ and the prior $p(f|\mathcal{X}, \boldsymbol{\theta}_{cov})$:

$$p(z|\mathcal{X}, \boldsymbol{\theta}) = \int p(z|f, \boldsymbol{\theta}_{lik}) \cdot p(f|\mathcal{X}, \boldsymbol{\theta}_{cov}) \, df \propto p(y|\mathcal{X}, \boldsymbol{\theta})$$

The last proportional sign holds if the warping function is fixed, meaning that it does
not contain hyper-parameters that need to be optimized.

Typically a ML-II or MAP-II optimization technique is applied. But in variations of the
20   present embodiment other optimization techniques may be applied as well.

When applying the ML-II method then the maximization of the marginal likelihood
with respect to the hyper parameters is achieved by minimizing the negative log
marginal likelihood with respect to the hyper parameters:

$$\boldsymbol{\theta}_{ML-II} = \underset{\boldsymbol{\theta}}{\mathrm{argmax}} \; p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}(-\log p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta}))$$

According to a variation of the present embodiment a method known as MAP-II may be applied, wherein the marginal likelihood is regularized with a suitable hyper prior distribution, $p(\theta)$, such as e.g. the half-student's-t distribution, the Gamma distribution, the Laplace distribution, the Gaussian distribution or a uniform prior for noise parameters.

Subsequently the regularized marginal likelihood is maximized with respect to $p(\boldsymbol{\theta}|\boldsymbol{z}, \mathcal{X}) \propto p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta})p(\boldsymbol{\theta})$ by finding the minimum of the negative logarithmic. Thus the MAP-II optimization is given by:

$$\boldsymbol{\theta}_{MAP-II} = \underset{\boldsymbol{\theta}}{\mathrm{argmax}} \; p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta})p(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}(-\log p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta}) - \log p(\boldsymbol{\theta}))$$

$$= \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \left(\frac{1}{2} \log \det(\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m}) + \frac{1}{2} \boldsymbol{z}^T (\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m})^{-1} \boldsymbol{z} - \log p(\boldsymbol{\theta}) \right),$$

The inventor has found that this method is especially advantageous when fewer than say 50 observations are available, which is typically the case when customizing hearing aid systems.

A new analytical expression for the marginal likelihood is derived by marginalizing the joint distribution between function values and observations over the function values. This joint distribution is the product of the likelihood and the prior hereby providing:

$$p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m})$$

When using the marginal likelihood for either MAP-II or ML-II optimization of hyper parameters it is typically more numerically robust to minimize the negative logarithmic of the marginal likelihood:

$$-\log p(\boldsymbol{z}|\mathcal{X}, \boldsymbol{\theta}) = \frac{m}{2} \log 2\pi + \frac{1}{2} \log \det(\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m}) + \frac{1}{2} \boldsymbol{z}^T (\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m})^{-1} \boldsymbol{z}$$

or if considering the original non-warped observations $\mathcal{Y}$:

$$-\log p(\boldsymbol{y}|\mathcal{X}, \boldsymbol{\theta}) = \frac{m}{2} \log 2\pi + \frac{1}{2} \log \det(\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m}) + \frac{1}{2} \boldsymbol{z}^T (\boldsymbol{MKM}^T + \sigma^2 \boldsymbol{I}_{m \times m})^{-1} \boldsymbol{z}$$

$$- \sum_{k=1}^{m} \log \frac{\partial g(y)}{\partial y} \bigg|_{y_k}$$

13

Note, that the last term appended for the negative log marginal likelihood of y is the Jacobian, and that the Jacobian does not depend on the hyper parameters, if the warping is fixed. Therefore, the Jacobian does not influence the gradient ascend/descend optimization of hyper parameters, and can therefore be neglected when performing the ML-II or MAP-II optimization of hyper parameters.

According to a variation of the present embodiment the warping function is the inverse cumulative density function of the Gaussian distribution, Φ^{-1} (y), in which case the Jacobian term is easily found as:

$$g(y) = \Phi^{-1}(y) \Rightarrow \frac{\partial g(y)}{\partial y} = \frac{1}{\mathcal{N}(\Phi^{-1}(y)|0,1)}$$

Thus the inventor has derived an analytical expression for the negative logarithmic of the marginal likelihood and hereby also the gradient of the negative logarithmic of the marginal likelihood with respect to the hyper parameters whereby the hyper parameters may be determined in a very processing efficient manner.

However, in variations of the present embodiment the hyper parameters of the covariance and likelihood may simply be set using experience from similar situations to provide a qualified guess. In a specific variation the hyper parameters are set based on experience from other hearing aid system users.

In other variations the warping function may contain hyper parameters.

In a tenth step an analytical posterior distribution over the unknown function values of the user's internal response function $p(f|z, \mathcal{X}, \theta)$ is derived based on Bayes rule:

$$p(f|z, \mathcal{X}, \theta) = \frac{p(z|f, \theta_{lik}) \cdot p(f|\mathcal{X}, \theta_{cov})}{p(z|\mathcal{X}, \theta)}$$

By applying the novel pairwise Gaussian likelihood given previously, the prior derived using a Gaussian process and the marginal likelihood as the integral over the user's internal response function of the product of the likelihood and the prior, then an analytical expression for the posterior can be found as:

$$p(f|z, \mathcal{X}, \theta) = \mathcal{N}(f|\mu, \Sigma)$$

$$\mu = K(M^T M K + \sigma^2 I_{n \times n})^{-1} M^T z$$

$$\Sigma = \left(K^{-1} + \frac{1}{\sigma^2} M^T M \right)^{-1}$$

14

In an eleventh step an analytical expression for the predictive distribution over the unknown function values of the user's internal response function is found. In full Bayesian modeling, predictions come in terms of a predictive distribution of new function values, $f^* = [f(x_1^*), \ldots, f(x_{n^*}^*)]^T$ given the observations, $\mathbf{y}$ or in case of warped observations $\mathbf{z}$. Hence, the predictive distribution is a conditional distribution and the conditional is on the observations only. It is derived from

$$p(f^*|\mathbf{z}, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}) = \int p(f^*|f, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}_{cov}) \, p(f|\mathbf{z}, \mathcal{X}, \boldsymbol{\theta}) \, df$$

Note, that $p(f^*|f, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}_{cov})$ is Gaussian due to the Gaussian Process as already discussed previously with reference to the eight step. Since the posterior distribution, according to the present embodiment is a Gaussian distribution on the form $p(f|\mathbf{z}, \mathcal{X}, \boldsymbol{\theta}) = \mathcal{N}(f|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the solution to the integral has an analytical solution and the predictive distribution is given, by inserting the mean and covariance from the posterior distribution, hereby obtaining:

$$p(f^*|\mathbf{z}, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}) = \int p(f^*|f, \mathcal{X}, \mathcal{X}^*, \boldsymbol{\theta}_{cov}) \, p(f|\mathbf{z}, \mathcal{X}, \boldsymbol{\theta}) \, df$$
$$= \int \mathcal{N}(f^*| K_*^T K^{-1} f, K_{**} - K_*^T K^{-1} K_*) \cdot \mathcal{N}(f| \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, df$$
$$= \mathcal{N}(f^*|K_*^T K^{-1} \boldsymbol{\mu}, K_{**} - K_*^T(K^{-1} + K^{-1}\boldsymbol{\Sigma}K^{-1})K_*^T)$$
$$= \mathcal{N}(f^*|K_*^T(M^T M K + \sigma^2 I_{n\times n})^{-1} M^T \mathbf{z}, K_{**} - K_*^T(M^T M K + \sigma^2 I_{n\times n})^{-1} M^T M K_*)$$

where:

$$K_* = \begin{bmatrix} k(x_1, x_1^*, \boldsymbol{\theta}_{cov}) & \cdots & k(x_1, x_{n^*}^*, \boldsymbol{\theta}_{cov}) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1^*, \boldsymbol{\theta}_{cov}) & \cdots & k(x_n, x_{n^*}^*, \boldsymbol{\theta}_{cov}) \end{bmatrix}$$

and:

$$K_{**} = \begin{bmatrix} k(x_1^*, x_1^*, \boldsymbol{\theta}_{cov}) & \cdots & k(x_1^*, x_{n^*}^*, \boldsymbol{\theta}_{cov}) \\ \vdots & \ddots & \vdots \\ k(x_{n^*}^*, x_1^*, \boldsymbol{\theta}_{cov}) & \cdots & k(x_{n^*}^*, x_{n^*}^*, \boldsymbol{\theta}_{cov}) \end{bmatrix}$$

Consider now the mean vector of the predictive distribution and denote it $\boldsymbol{\mu}^*$, which from the expressions above is given by

$$\boldsymbol{\mu}^* = K_*^T(M^T M K + \sigma^2 I_{n\times n})^{-1} M^T \mathbf{z}$$

wherefrom it directly follows that only the term $\boldsymbol{K}_*$ depends on the parameter value settings belonging to $\mathcal{X}^*$.

In a twelfth step the parameter value setting, that the user prefers among all possible settings (i.e. the set $\mathcal{X}^*$) can be found by considering a case where $\mathcal{X}^*$ contains only one setting, $\boldsymbol{x}^*$, implying that $n^* = 1$, whereby we get:

$$\boldsymbol{K}_* = [k(\boldsymbol{x}^*, \boldsymbol{x}_1, \boldsymbol{\theta}_{cov}), \dots, k(\boldsymbol{x}^*, \boldsymbol{x}_n, \boldsymbol{\theta}_{cov})]^T = \boldsymbol{k}(\boldsymbol{x}^*)$$

This is a function, which takes a single input $\boldsymbol{x}^*$ and returns a vector of covariance function outputs for every tested parameter value setting, $\boldsymbol{x}_i$. Thereby, we can interpret the mean value of the predictive distribution as a function of a single parameter value setting that returns the corresponding mean function value as

$$\mu^*(\boldsymbol{x}^*) = \boldsymbol{k}(\boldsymbol{x}^*)^T (\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{K} + \sigma^2 \boldsymbol{I}_{n \times n})^{-1} \boldsymbol{M}^T \boldsymbol{z} = \boldsymbol{k}(\boldsymbol{x}^*)^T \cdot \boldsymbol{B},$$

where the matrix B does not depend on $\boldsymbol{x}^*$.

This interpretation of the mean value as function taking a single input makes it possible to find the (local) maximum of the mean of the predictive distribution with respect to the input, $\boldsymbol{x}^*$, using the gradient given by:

$$\frac{\partial \mu^*(\boldsymbol{x}^*)}{\partial \boldsymbol{x}^*} = \left[\frac{\partial \boldsymbol{k}_*(\boldsymbol{x}^*)}{\partial \boldsymbol{x}^*} \right]^T (\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{K} + \sigma^2 \boldsymbol{I}_{n \times n})^{-1} \boldsymbol{M}^T \boldsymbol{z}.$$

Thus it follows that the present invention offers an estimate of the parameter value setting that the user prefers, among all the possible parameter value settings, wherein the estimate can be provided using only very limited processing and memory resources due to the availability of an analytical expression for the gradient.

However, according to a variation of the present embodiment the parameter value setting that the user prefers, among the tested settings, can be found by considering that in this case the parameter value settings in the set $\mathcal{X}^*$ are the same as the settings that have been presented to the user (i.e. the settings belonging to $\mathcal{X}$) and consequently we have:

$$\boldsymbol{K}_* = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1, \boldsymbol{\theta}_{cov}) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_n, \boldsymbol{\theta}_{cov}) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1, \boldsymbol{\theta}_{cov}) & \cdots & k(\boldsymbol{x}_n, \boldsymbol{x}_n, \boldsymbol{\theta}_{cov}) \end{bmatrix} = \boldsymbol{K}$$

and consequently:

$$\boldsymbol{\mu}^* = \boldsymbol{K}_*^T(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{z} = \boldsymbol{K}(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{z}$$

Hereby an estimate of the parameter value setting that the user prefers can be found with a method that requires even fewer processing and memory resources than the present embodiment, because the set $\mathcal{X}$ is typically of limited size and therefore can be found without a gradient approach.

5   However, according to the present embodiment, the settings that the user is prompted to compare are not selected randomly. Instead the next new setting, to be compared with the current best parameter value setting $\hat{\boldsymbol{x}}$, is found as the parameter value setting $\hat{\boldsymbol{x}}^*$ that maximizes a new bivariate Expected Improvement, given by:

$$\hat{\boldsymbol{x}}^* = \underset{\boldsymbol{x}^*}{\operatorname{argmax}}\left(\mu_I\Phi\left(\frac{\mu_I}{\sigma_I}\right) + \sigma_I\mathcal{N}\left(\frac{\mu_I}{\sigma_I}\middle|0,1\right)\right),$$

wherein:

$$\mu_I = \left(\boldsymbol{k}_*(\boldsymbol{x}^*) - \boldsymbol{k}_*(\hat{\boldsymbol{x}})\right)\cdot(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{z},$$

$$\sigma_I^2 = k(\boldsymbol{x}^*,\boldsymbol{x}^*) + k(\hat{\boldsymbol{x}},\hat{\boldsymbol{x}}) - k(\boldsymbol{x}^*,\hat{\boldsymbol{x}}) - \boldsymbol{k}_*(\boldsymbol{x}^*)^T(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{k}_*(\boldsymbol{x}^*)$$

$$- \boldsymbol{k}_*(\hat{\boldsymbol{x}})^T(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{k}_*(\hat{\boldsymbol{x}})$$

$$- \boldsymbol{k}_*(\boldsymbol{x}^*)^T(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{k}_*(\hat{\boldsymbol{x}}),$$

10 and:

$$\boldsymbol{k}_*(\boldsymbol{x}) = [k(\boldsymbol{x},\boldsymbol{x}_i),\dots,k(\boldsymbol{x},\boldsymbol{x}_n)]$$

and wherein $\Phi\left(\frac{\mu_I}{\sigma_I}\right)$ is the standard cumulative distribution function of the Gaussian distribution.

It is a further specific advantage of the analytical expression for the bivariate Expected Improvement that it allows an analytical expression for the gradient of the bivariate

15 Expected improvement to be derived:

$$\frac{\partial}{\partial\boldsymbol{x}^*}\left\{\mu_I\Phi\left(\frac{\mu_I}{\sigma_I}\right) + \sigma_I\mathcal{N}\left(\frac{\mu_I}{\sigma_I}\middle|0,1\right)\right\} = \frac{\partial\sigma_I}{\partial\boldsymbol{x}^*}\mathcal{N}\left(\frac{\mu_I}{\sigma_I}\middle|0,1\right) + \frac{\partial\mu_I}{\partial\boldsymbol{x}^*}\Phi\left(\frac{\mu_I}{\sigma_I}\right),$$

Where

$$\frac{\partial\mu_I}{\partial\boldsymbol{x}^*} = \left[\frac{\partial\boldsymbol{k}_*(\boldsymbol{x}^*)}{\partial\boldsymbol{x}^*}\right]^T(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{K} + \sigma^2\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{M}^T\boldsymbol{z}$$

and

$$\frac{\partial \sigma_I}{\partial \boldsymbol{x}^*} = \frac{1}{\sigma_I} \left(\frac{1}{2} \frac{\partial k(\boldsymbol{x}^*, \boldsymbol{x}^*)}{\partial \boldsymbol{x}^*} - \left[\frac{\partial \boldsymbol{k}_*(\boldsymbol{x}^*)}{\partial \boldsymbol{x}^*} \right]^T (\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{K} + \sigma^2 \boldsymbol{I}_{n \times n})^{-1} \boldsymbol{M}^T \boldsymbol{M} \boldsymbol{k}_*(\boldsymbol{x}^*) - \frac{\partial k(\boldsymbol{x}^*, \hat{\boldsymbol{x}})}{\partial \boldsymbol{x}^*} \right.$$

$$\left. + \left[\frac{\partial \boldsymbol{k}_*(\boldsymbol{x}^*)}{\partial \boldsymbol{x}^*} \right]^T (\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{K} + \sigma^2 \boldsymbol{I}_{n \times n})^{-1} \boldsymbol{M}^T \boldsymbol{M} \boldsymbol{k}_*(\hat{\boldsymbol{x}}) \right)$$

When deriving the new analytical expression for the bivariate Expected Improvement the inventor has considered that when using a Gaussian Process to model the function value, $f(\boldsymbol{x}^*)$, of a new (i.e. not yet tested) parameter value setting $\boldsymbol{x}^*$, then the function value of the new parameter value setting, $\boldsymbol{x}^*$, co-varies with the function value, $f(\hat{\boldsymbol{x}})$, at the current maximum, $\hat{\boldsymbol{x}}$. The original and well known (uni-variate) Expected Improvement formulation (which in the following may be referred to as simply EI) does not consider this. Instead, the EI formulation requires the, $f(\hat{\boldsymbol{x}})$, to be deterministically defined. Consequently, when using EI with Gaussian Processes the mean function value at the maximum point $\hat{\mu}$ is used as the deterministic prediction of the corresponding function value, $f(\hat{\boldsymbol{x}})$, thus neglecting both the variance, $\hat{\sigma}^2$, and covariance with the function value for the new parameter value setting available from the predictive distribution from the eleventh step. In a very common scenario, this has the undesirable consequence that among all possible new parameter value settings, $\boldsymbol{x}^*$, the parameter value setting with the largest EI is the setting that is arbitrarily close to the current maximum, $\hat{\boldsymbol{x}}$. Thereby, the EI criterion will end up suggesting points arbitrarily close to the current maximum, only because the covariance is neglected. The bivariate Expected Improvement designed by the inventor avoids this undesirable behavior, because the covariance is included.

On the downside, the bivariate Expected Improvement approach requires the entire covariance matrix of the predictive distribution, from the eleventh step, to be computed and stored. With only a small number of parameters, say more than three, the size of the covariance matrix will be way too big to be stored in a hearing aid system memory, even with tens of GB of available memory. Therefore it is of significant importance that the inventor has provided an analytical expression for the bivariate Expected Improvement because this allows it to be maximized with a gradient ascend procedure, which does not require the entire covariance matrix to be computed and stored.

18

In variations of the first embodiment the next new setting needs not be determined using the new and inventive bivariate Expected Improvement method. Instead the next new setting may be determined using a single variate Expected Improvement method.

In other variations of the first embodiment the inventor has found that any customization of hearing aid system settings may benefit from the disclosed method.

Reference is now made to a method of customizing parameters in a hearing aid system according to a second embodiment of the invention.

In a first step a multitude of hearing aid system parameters are selected for user customization. According to the present embodiment the hearing aid system comprises a graphical user interface that allows the user to select the parameters to be customized. In variations of the present embodiment the hearing aid system is adapted such that the user is offered to select from a menu whether to customize e.g. noise reduction or relaxing sounds and the hearing aid system will in response to the user's selection select appropriate parameters to be customized and the ranges within which the parameters are allowed to vary.

In a second step the user is prompted to compare the acoustical output from the hearing aid system as a function of two different parameter value settings and provide a user response representing the user's rating of the two different parameter value settings relative to each other. According to the present embodiment the hearing aid system comprises a graphical user interface that allows the user to input his rating.

In a third step a multitude of user responses and corresponding parameter value settings are stored, at least temporary in a memory of the hearing aid system and subsequently the parameter value settings used to determine a covariance matrix K, of size n x n wherein n is the number of tested parameter value settings given as:

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

wherein k is a positive semi-definite covariance function such as the squared exponential covariance function, and x_1, x_2,..., x_d are vectors representing different parameter value settings.

In a fourth step a matrix **M** is provided, of size m x m, where m is the number of provided user responses, wherein **M** comprises only zero's except for $[M]_{k,u_k} = 1$ and $[M]_{k,v_k} = -1$, and wherein index k represents the k'th user response and indices u_k and v_k represents the parameter value settings used to prompt the k'th user response.

According to a variation of the present embodiment the parameter value setting, that the user prefers $\hat{x}$, among the tested parameter value settings, is given as:

$$\hat{x} = \underset{x}{\mathrm{argmax}}\ K^T(M^TMK + \sigma^2 I_{n\times n})^{-1}M^Tz$$

wherein σ^2 is Gaussian noise that is independent and identically distributed and represents the uncertainty of the user when carrying out the graduated responses and wherein **z** is a vector derived from the provided user responses.

According to the present embodiment σ^2 is determined based on a MAP-II optimization as explained in detail for the first embodiment of the present invention and the vector **z** is derived from the vector **y** of bounded user responses, also as explained in detail for the first embodiment of the present invention.

Reference is now made to Fig. 1, which illustrates highly schematically a hearing aid system 100 according to a third embodiment of the invention. The hearing aid system 100 comprises a hearing aid 101 and an external device 102. The external device 102 comprises a graphical user interface 103, a parameter setting selector 104, a parameter memory 105 and an optimum parameter estimator 106. The hearing aid 101 comprises an audio input 107, a hearing aid digital signal processor (DSP) 108, a parameter controller 109 and an electrical-acoustical output transducer 110.

Firstly the graphical user interface 103 is adapted to allow a hearing aid system user 111 to select a number of hearing aid parameters for customization to the hearing aid system user's preference. The parameter memory 105 holds information, on the parameters that may be selected for customization, such as the ranges wherein the parameters are allowed to vary.

The parameter setting selector 104 comprises an algorithm that allows the next two parameter value settings that are to be rated by the hearing aid user 111 to be determined, and the parameter setting selector 104 is further adapted to provide said

two parameter value settings to be transmitted to the parameter controller 109 of the hearing aid 101.

The parameter controller 109 is adapted to control either the audio input 107, in case sound is to be generated synthetically in the hearing aid or the hearing aid digital signal processor 108, in case the hearing aid DSP 108 uses the parameters to be rated when processing sound from the audio input 107.

The audio input 107 may either provide synthetically generated electrical signals representing e.g. relaxing sounds or may relay signals received from one or more acoustical-electrical transducers.

The hearing aid DSP 108 is adapted to process the electrical signals representing sounds that are received from the audio input 107 and provide the processed signals to the electrical-acoustical transducer 110, in order to alleviate a hearing loss by amplifying sound at frequencies in those parts of the audible frequency range where the user suffers a hearing deficit.

The optimum parameter estimator 106 is adapted to estimate the parameter value setting that the hearing aid system user 111 prefers based on the user response provided by the hearing aid system user 111 and the parameter value settings evaluated as described in great detail with reference to the previous embodiments. The optimum parameter estimator 106 is furthermore adapted to provide the preferred parameter setting to the hearing aid, in order to adjust the parameter setting in the hearing aid and hereby finalizing the customization process, This may be done in response to a user input triggering this, in response to the user having carried out a predetermined number of ratings or in response to some other convergence criteria being fulfilled.

According to variations of the previous embodiments the customization process is deemed to have converged when the combined magnitude of the Expected Improvement for a plurality of at least pseudo-randomly selected parameter value settings are below a predetermined threshold.

Generally the variations, mentioned in connection with a specific embodiment, may, where applicable, be considered variations for the other disclosed embodiments as well.

This is especially true with respect to the fact that the disclosed methods may be used for any type of optimization (i.e. customization) of hearing aid system settings.

This is also true with respect to the methods of selecting the values of the hyper parameters for both the covariance and likelihood functions.

5   Furthermore this is true with respect to whether mapping (that may also be denoted warping) techniques are applied.

This is likewise true for the methods for determining the next parameter value settings to be rated by the user. The present invention does not depend on a specific method for determining the next parameter value setting, although the new and inventive bivariate

10   Expected Improvement method may be significantly advantageous.

It is likewise independent on a specific embodiment whether the parameters to be customized are used to control how sound is processed in the hearing aid system or whether they are used to control how sound is synthetically generated by the hearing aid system.

15   Thus e.g. how the hearing aid system parameters are provided or offered or selected for user customization does not depend on a specific embodiment. Neither does the method of providing the user response depend on a specific embodiment.

In the present context the distributions over the unknown function values of the user's internal response function are primarily expressed such that the hyper parameters

20   appear explicitly. However, in case this is not the case everywhere the hyper parameters are obviously assumed to be implicitly disclosed.

CLAIMS

1. A method of optimizing parameters in a hearing aid system comprising the steps of:

   - providing a set of parameters x_1, x_2,..., x_d to be optimized;

   - providing a first sound based on a first parameter value setting $\mathbf{x_1} = [x_{11}, x_{21},..., x_{d1}]$ and providing a second sound based on a second parameter value setting $\mathbf{x_2} = [x_{12}, x_{22},..., x_{d2}]$;

   - prompting a user to rate said first and second sounds relative to each other;

   - providing a user response y_1 representing the user's rating of the two sounds relative to each other;

   - providing m user responses $\mathbf{y} = [y_1, y_2, ..., y_m]$ wherein each of the user responses represent the user's rating of two sounds relative to each other and wherein the sounds are derived from a multitude of n parameter value settings $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$;

   - defining a Gaussian likelihood function as:

   $$p(y_k | f(\boldsymbol{x}_u), f(\boldsymbol{x}_v), \sigma^2) = \mathcal{N}(y_k | f(\boldsymbol{x}_u) - f(\boldsymbol{x}_v), \sigma^2)$$

   wherein $f(\boldsymbol{x}_u)$ and $f(\boldsymbol{x}_v)$ represent specific function values of the user's internal response function f, wherein y_k represents a specific user response, wherein σ^2 represents the variance of the user response and wherein $\mathcal{N}(y_k | f(\boldsymbol{x}_u) - f(\boldsymbol{x}_v), \sigma^2)$ is the single variate Gaussian distribution over the variable y_k with mean value $f(\boldsymbol{x}_u) - f(\boldsymbol{x}_v)$ and variance σ^2;

   - defining the likelihood as:

   $$p(\boldsymbol{y} | \boldsymbol{f}) = \prod_{k=1}^{m} \mathcal{N}(y_k | f(\boldsymbol{x}_u) - f(\boldsymbol{x}_v), \sigma^2) = \mathcal{N}(\boldsymbol{y} | \boldsymbol{Mf}, \sigma^2 \boldsymbol{I})$$

   wherein $\mathcal{N}(\boldsymbol{y} | \boldsymbol{Mf}, \sigma^2 \boldsymbol{I}_{m \times m})$ represents a multivariate Gaussian distribution over the user responses $\boldsymbol{y}$ with mean vector $\boldsymbol{Mf}$ and covariance matrix $\sigma^2 \boldsymbol{I}$, wherein $\boldsymbol{I}$ is an m x m identity matrix, and $\boldsymbol{M}$ is a m x n matrix having only zero elements except for the elements $[\boldsymbol{M}]_{k,u_k} = 1$ and $[\boldsymbol{M}]_{k,v_k} = -1$;

   - obtaining a prior distribution over the function values of the user's internal response function using a Gaussian process;

   - using Bayes rule to obtain an analytical posterior distribution over the function values of the user's internal response function;

   - deriving an analytical expression for the predictive distribution based on the

analytical expression of the posterior distribution over the function values of the user's internal response function; and

- using the analytical expression for the predictive distribution to find the parameter value settings that the user prefers.

2. The method according to claim 1, wherein the user response takes the form of a number from a bounded or partly bounded interval.

3. The method according to any one of the preceding claims, wherein the user response is modelled in accordance with $y_k = f(\mathbf{x}_{u_k}) - f(\mathbf{x}_{v_k}) + \varepsilon$, wherein y_k is the user response, $f(\mathbf{x}_{u_k})$ and $f(\mathbf{x}_{v_k})$ are function values of the internal response function of the user for two different parameter value settings $\mathbf{x}_{u_k}$ and $\mathbf{x}_{u_k}$ and wherein ε is assumed to be identically and independently distributed Gaussian noise.

4. The method according to any one of the preceding claims, wherein said prior distribution over the function values of the user's internal response function is given as: $p(\mathbf{f}|\mathcal{X}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K})$, wherein $\mathcal{X}$ is a set of the n tested parameter value settings and wherein $\mathbf{K}$ is a covariance matrix.

5. The method according to claim 4, wherein the covariance matrix $\mathbf{K}$ is determined from: $[\mathbf{K}]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$, wherein k is a positive semi-definite covariance function selected from a group of functions comprising the squared exponential covariance function, the γ--exponential covariance function, the Matern class of covariance functions, the rational quadratic covariance function, periodic covariance functions and linear covariance functions.

6. The method according to any one of the claims 4 - 5, wherein the analytical expression for the posterior distribution over the function values of the user's internal response function is given as:

$$p(\mathbf{f}|\mathbf{y}, \mathcal{X}) = \mathcal{N}(\mathbf{f}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = \mathbf{K}(\mathbf{M}^T\mathbf{M}\mathbf{K} + \sigma^2\mathbf{I}_{n\times n})^{-1}\mathbf{M}^T\mathbf{y}$$

$$\boldsymbol{\Sigma} = \left(\mathbf{K}^{-1} + \frac{1}{\sigma^2}\mathbf{M}^T\mathbf{M}\right)^{-1}$$

7. The method according to any one of the preceding claims, wherein the analytical expression for the predictive distribution is given as:

$$\mathcal{N}\left(f^* | K_*^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T y, K_{**} - K_*^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M K_*\right)$$

where:

$$K_* = \begin{bmatrix} k(x_1, x_1^*, \theta_{cov}) & \cdots & k(x_1, x_{n^*}^*, \theta_{cov}) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1^*, \theta_{cov}) & \cdots & k(x_n, x_{n^*}^*, \theta_{cov}) \end{bmatrix}$$

and:

$$K_{**} = \begin{bmatrix} k(x_1^*, x_1^*, \theta_{cov}) & \cdots & k(x_1^*, x_{n^*}^*, \theta_{cov}) \\ \vdots & \ddots & \vdots \\ k(x_{n^*}^*, x_1^*, \theta_{cov}) & \cdots & k(x_{n^*}^*, x_{n^*}^*, \theta_{cov}) \end{bmatrix}$$

wherein θ_{cov} and θ_{lik} are hyper parameters for the covariance and likelihood.

8. The method according to any one of the preceding claims, wherein the step of providing a set of user responses comprises the step of determining the next parameter value settings to be rated by the user using an analytical expression for the bivariate expected improvement.

9. The method according to claim 8, wherein the analytical expression for the bivariate expected improvement is given by:

$$\mu_I \Phi\left(\frac{\mu_I}{\sigma_I}\right) + \sigma_I \mathcal{N}\left(\frac{\mu_I}{\sigma_I} \middle| 0,1\right)$$

wherein:

$$\mu_I = \left(k_*(x^*) - k_*(\hat{x})\right) \cdot (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T y$$

wherein:

$$\mu_I = \left(k_*(x^*) - k_*(\hat{x})\right) \cdot (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T y,$$

$$\sigma_I^2 = k(x^*, x^*) + k(\hat{x}, \hat{x}) - k(x^*, \hat{x}) - k_*(x^*)^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M k_*(x^*)$$
$$- k_*(\hat{x})^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M k_*(\hat{x})$$
$$- k_*(x^*)^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M k_*(\hat{x})$$

and:

$$k_*(x) = [k(x, x_i), \dots, k(x, x_n)]$$

and wherein $\Phi\left(\frac{\mu_I}{\sigma_I}\right)$ is the standard Gaussian cumulative distribution function.

10. The method according to claim 8 or 9, wherein an analytical expression for the gradient of the bivariate expected improvement as a function of a new parameter

value setting is used to find a maximum of the bivariate expected improvement as a function of the set of parameter values using a gradient ascend or descend approach.

11. The method according to claim 10, wherein the gradient of the analytical expression for the gradient of the bivariate expected improvement as a function of the set of parameter values is given as:

$$\frac{\partial}{\partial x^*}\left\{\mu_I \Phi\left(\frac{\mu_I}{\sigma_I}\right) + \sigma_I \mathcal{N}\left(\frac{\mu_I}{\sigma_I}\middle| 0,1\right)\right\} = \frac{\partial \sigma_I}{\partial x^*}\mathcal{N}\left(\frac{\mu_I}{\sigma_I}\middle| 0,1\right) + \frac{\partial \mu_I}{\partial x^*}\Phi\left(\frac{\mu_I}{\sigma_I}\right),$$

where:

$$\frac{\partial \mu_I}{\partial x^*} = \left[\frac{\partial k_*(x^*)}{\partial x^*}\right]^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T y$$

and:

$$\frac{\partial \sigma_I}{\partial x^*} = \frac{1}{\sigma_I}\left(\frac{1}{2}\frac{\partial k(x^*, x^*)}{\partial x^*} - \left[\frac{\partial k_*(x^*)}{\partial x^*}\right]^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M k_*(x^*) - \frac{\partial k(x^*, \hat{x})}{\partial x^*}\right.$$
$$\left. + \left[\frac{\partial k_*(x^*)}{\partial x^*}\right]^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T M k_*(\hat{x})\right)$$

12. The method according to any one of the preceding claims, comprising the steps of: determining covariance and likelihood hyper parameters θ_{cov} and θ_{lik} by applying a ML-II or MAP-II optimization on the marginal likelihood.

13. The method according to claim 12, wherein the step of determining the likelihood and covariance hyper parameters comprises finding the minimum of the negative logarithmic regularized marginal likelihood, as a function of the hyper parameters, given by:

$$\underset{\theta}{\operatorname{argmin}}\left(\frac{1}{2}\log\det(MKM^T + \sigma^2 I_{m \times m}) + \frac{1}{2}y^T(MKM^T + \sigma^2 I_{m \times m})^{-1}y - \log p(\theta)\right)$$

wherein p(θ) is a suitable hyper prior distribution.

14. The method according to any one of the preceding claims, wherein the parameters to be optimized comprises parameters for controlling the generation of relaxing sounds.

15. The method according to any one of the preceding claims, comprising the step of: mapping the user responses from a first response range that is bounded or partly bounded and to a second response range that is unbounded.

16. The method according to claim 15, wherein the step of mapping the set of user responses is adapted such that the user responses becomes more Gaussianly distributed in the second response range.

17. The method according to claim 15 or 16, wherein the step of mapping the user responses is carried out using a monotonically increasing function selected from a group of functions comprising the inverse cumulative distribution function of the Gaussian distribution, inverse sigmoid function and inverse hyperbolic tangent function.

18. A method of optimizing parameters in a hearing aid system comprising the steps of:
   - providing a set of parameters x_1, x_2,..., x_d to be optimized;
   - providing a first sound based on a first parameter value setting $\mathbf{x_1} = [x_{11}, x_{21},.., x_{d1}]$ and providing a second sound based on a second parameter value setting $\mathbf{x_2} = [x_{12}, x_{22},..., x_{d2}]$;
   - prompting a user to rate said first and second sounds relative to each other;
   - providing a user response y_1 representing the user's rating of the two sounds relative to each other;
   - providing m user responses $\mathbf{y} = [y_1, y_2, ..., y_m]$ wherein each of the user responses represent the user's rating of two sounds relative to each other and wherein the sounds are derived from a multitude of n parameter value settings $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_n}$;
   - obtaining a matrix $\mathbf{K}$ as:

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x_1}, \mathbf{x_1}) & \cdots & k(\mathbf{x_1}, \mathbf{x_n}) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x_n}, \mathbf{x_1}) & \cdots & k(\mathbf{x_n}, \mathbf{x_n}) \end{bmatrix}$$

wherein k is a positive semi-definite covariance function,
   - obtaining a matrix $\mathbf{M}$, that is a m x n matrix comprising only zero's except for: $[\mathbf{M}]_{k,u_k} = 1$ and $[\mathbf{M}]_{k,v_k} = -1$;
   - determining the parameter value setting, that the user prefers $\hat{\mathbf{x}}$, among the tested parameter value settings as:

$$\hat{\mathbf{x}} = \underset{x}{\text{argmax }} \mathbf{K}^T (\mathbf{M}^T \mathbf{M} \mathbf{K} + \sigma^2 \mathbf{I}_{n \times n})^{-1} \mathbf{M}^T \mathbf{y}$$

and wherein σ^2 is Gaussian noise that is independent and identically distributed and represents the uncertainty of the user when carrying out the graduated responses.

19. A method of optimizing parameters in a hearing aid system comprising the steps of:

- providing a set of parameters x_1, x_2,..., x_d to be optimized;

- providing a first sound based on a first parameter value setting $\mathbf{x_1} = [x_{11}, x_{21},..., x_{d1}]$ and providing a second sound based on a second parameter value setting $\mathbf{x_2} = [x_{12},$

5      $x_{22},..., x_{d2}]$;

- prompting a user to rate said first and second sounds relative to each other;

- providing a user response y_1 representing the user's rating of the two sounds relative to each other;

- providing m user responses $\mathbf{y} = [y_1, y_2, ..., y_m]$ wherein each of the user responses

10   represent the user's rating of two sounds relative to each other and wherein the sounds are derived from a multitude of n parameter value settings $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_n}$;

- obtaining a matrix $\mathbf{K}$ as:

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

wherein k is a positive semi-definite covariance function,

- obtaining a matrix $\mathbf{M}$, that is a m x n matrix comprising only zero's except for:

15     $[\mathbf{M}]_{k,u_k} = 1$ and $[\mathbf{M}]_{k,v_k} = -1$;

- determining the parameter value setting, that the user prefers, among all possible parameter value settings $\mathbf{x}^*$ using the mean value of the predictive distribution $\mu^*(\mathbf{x}^*)$ as function of a single parameter value setting:

$$\mu^*(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T (\mathbf{M}^T \mathbf{M} \mathbf{K} + \sigma^2 \mathbf{I}_{n \times n})^{-1} \mathbf{M}^T \mathbf{y}$$

wherein

$$\mathbf{k}(\mathbf{x}^*) = [k(\mathbf{x}^*, \mathbf{x}_1, \boldsymbol{\theta}_{cov}), ..., k(\mathbf{x}^*, \mathbf{x}_n, \boldsymbol{\theta}_{cov})]^T$$

20   wherein θ_{cov} and θ_{lik} are hyper parameters for the covariance and likelihood functions and wherein σ^2 is Gaussian noise that is independent and identically distributed and represents the uncertainty of the user when carrying out the graduated responses;

- using a gradient ascent or gradient descent approach to find a maximum of the

25   mean of the predictive distribution with respect to the input, $\mathbf{x}^*$, using the gradient given by:

$$\frac{\partial \mu^*(x^*)}{\partial x^*} = \left[\frac{\partial k_*(x^*)}{\partial x^*}\right]^T (M^T M K + \sigma^2 I_{n \times n})^{-1} M^T y.$$

20. A hearing aid system adapted to optimize parameters in a hearing aid system in accordance with a method according to any one of the preceding claims.

21. A hearing aid system according to claim 20, wherein the parameters that the hearing aid system is adapted to optimize comprise parameters for controlling the generation of relaxing sounds.
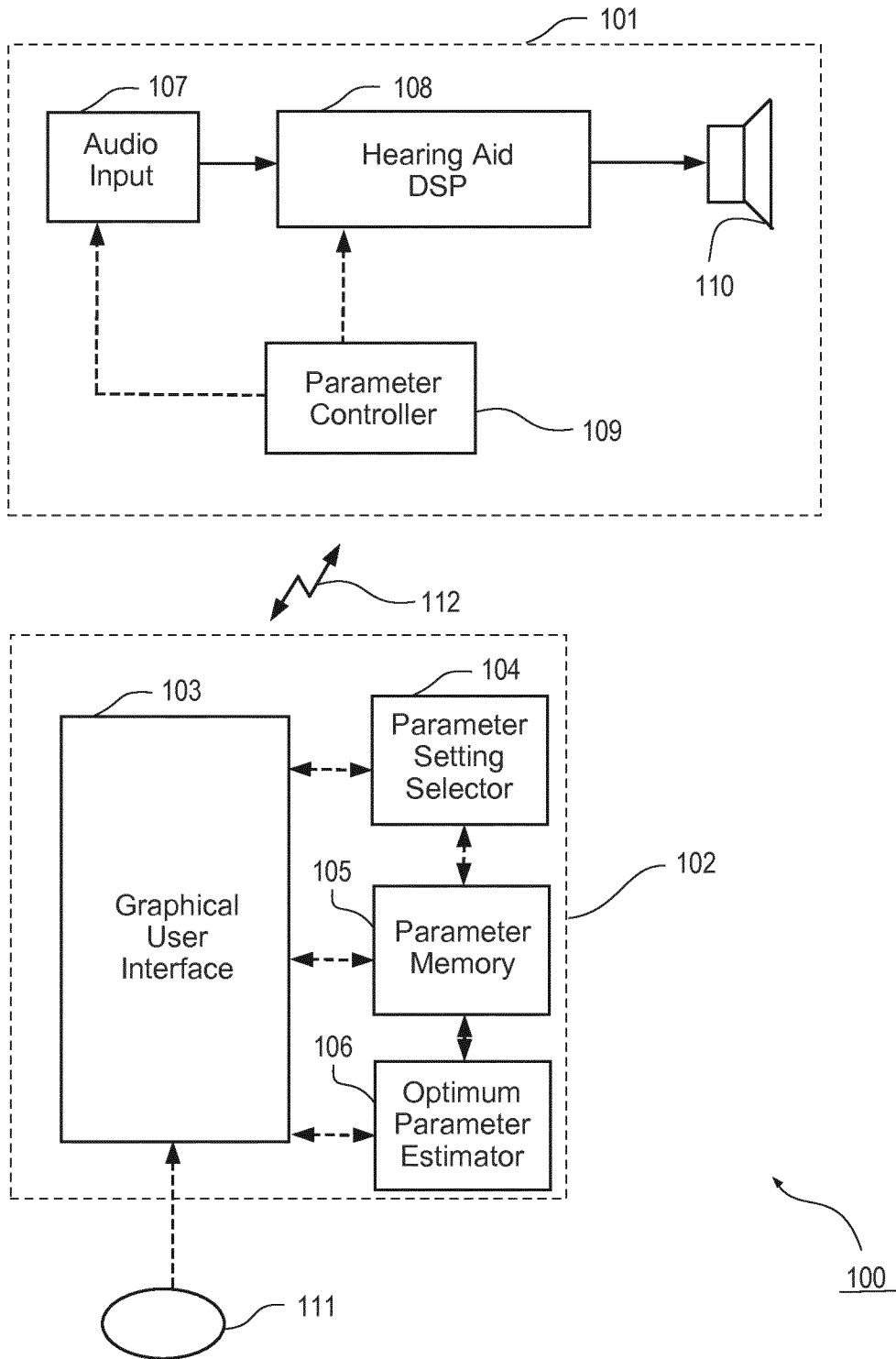
Fig. 1

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04R25/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2007/042043 A2 (GN RESOUND AS [DK]; DE VRIES AALBERT [NL]; YPMA ALEXANDER [NL]) 19 April 2007 (2007-04-19) cited in the application page 2, lines 16-19 page 6, lines 1-5 page 8, lines 24-27 page 13, lines 11-23 page 22, line 23 - page 23, line 21 ----- -/--	1-21

X	Further documents are listed in the continuation of Box C.		X	See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 February 2015	05/03/2015

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Rogala, Tomasz

1

Form PCT/ISA/210 (second sheet) (April 2005)

page 1 of 2

INTERNATIONAL SEARCH REPORT

C(Continuation).   DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ALEXANDER YPMA ET AL: "Bayesian Feature Selection for Hearing Aid Personalization", MACHINE LEARNING FOR SIGNAL PROCESSING, 2007 IEEE WORKSHOP ON, IEEE, PI, 1 August 2007 (2007-08-01), pages 425-430, XP031199124, ISBN: 978-1-4244-1565-6 section 3.1; page 427	1-21
A	GROOT P C ET AL: "Predicting Preference Judgments of Individual Normal and Hearing-Impaired Listeners With Gaussian Processes", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 19, no. 4, 1 May 2011 (2011-05-01), pages 811-821, XP011352007, ISSN: 1558-7916, DOI: 10.1109/TASL.2010.2064311 section D; page 814	1-21

1

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 2007042043	A2	19-04-2007	AT	469514	T	15-06-2010
			EP	1946609	A2	23-07-2008
			US	2010008526	A1	14-01-2010
			WO	2007042043	A2	19-04-2007