Evidence-based hearing aid algorithms

Bert de Vries[1,2], Tjeerd Dijkstra[1,2], Alexander Ypma[1] and Jos Leenen[1]

[1] *Algorithm R&D, GN ReSound, Eindhoven, The Netherlands, (www.bertdv.nl)*

[2] *Signal Processing Systems group, dept. E.E., Technical University Eindhoven, The Netherlands*

Hearing aid (HA) algorithms contain a large number of tuning parameters that need to be optimized with respect to the expected patient satisfaction. Here we report on a new fitting-engineering approach where patient measurements (audiogram, listening tests etc.) are transferred without loss of information to optimal values for HA tuning parameters. Our approach is rooted in Bayesian decision theory and takes properly account of inconsistencies in the measured patient data. The presented approach is envisioned to assist the experienced HA dispenser in the challenging task of fitting a HA algorithm to a specific patient in a limited time.

INTRODUCTION

Consider a HA algorithm $y = H(x,\theta)$, where   and   are acoustic input and output signals, respectively, and θ is a vector of HA *tuning parameters* (like compression ratios, time constants, filter coefficients, etc.). There is an essential difference between the arguments x and θ. While it is easy to record x with a microphone, it is very hard to get good values for θ. After all, we cannot measure tuning parameter values in nature; instead, we have to choose appropriate values for θ ourselves.

In order to appreciate the complexity of this problem, consider the following back-of-the-envelope calculation. A typical HA algorithm contains about 140 tuning parameters (say, 15 frequency bands times 7 parameters shared by the AGC and spectral subtraction modules, plus 35 filter taps shared between the feedback cancellation and beamforming filters). If we assume that each parameter can take on 5 interesting values (very low, low, medium, high, very high), then the total number of potentially interesting algorithm configurations is 5^{140}. This is far more than 5^{115}, the number of electrons in the universe. Hence, at face value, finding the optimal parameter values for a specific patient (the fitting task) appears to be at least as complex as finding a specific electron in the universe.

How do we deal with this complexity? We must somehow determine good values for θ based on measurements about a patient. What can we know about this patient? We can measure his auditory profile a, which refers to a set of variables including the audiogram, SNR loss and demographic data. Moreover, we can carry out listening tests (e.g. paired comparison tests) with this patient and collect all results in a multidimensional variable D. The data set $\{D,a\}$ contains all information that we can acquire about a patient.

Rather than an algorithm $y = H(x,\theta)$, we would therefore prefer to have access to an

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

alternative algorithm *y=H'(x,D,a)*, because in the latter case all arguments (*x,D*, and *a*) can be measured in nature. Hence, if we have access to algorithm *H'* (H-prime), we are not challenged to choose values for the arguments, but instead just measure them, and apply the algorithm *H'(x,D,a)* instead of *H(x,θ)*. Since *H'* takes as its arguments all available data (evidence) about the problem at hand, we call *H'(x,D,a)* an *evidence-based* hearing aid algorithm, in contrast to *H(x,θ)*, which is 'just' a HA algorithm.

We have developed a consistent mathematical theory for designing evidence-based hearing aid algorithms from a given HA algorithm *H* and a measured data set *{D,a}*. In other words, we have developed a mathematical theory for fitting hearing aids. In this paper, we will describe our theory for evidence-based hearing aids in (relatively) easy mathematical terms.

HEARING AID TUNING THROUGH EXPECTED UTILITY MAXIMIZATION

For a given patient, our objective is to choose values for the tuning parameters θ such that the 'expected patient satisfaction' is maximized. Clearly, we cannot measure patient satisfaction directly, since we can only measure *{D,a}*. Let us introduce a model for patient satisfaction, called utility model (name is derived from decision theory),

$$z = U(y) + \varepsilon_z \qquad \text{(Eq. 1)}$$

where z is the listener's perceived utility (the perceived satisfaction rate) for signal y, $U(y)$ a deterministic utility model and ε_z is an internal noise term that captures the part of that cannot be modelled by $U(y)$, see Figure 1.

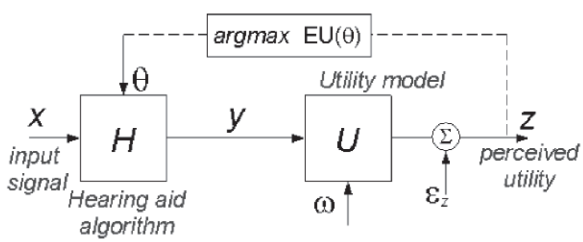

Fig. 1. Algorithm tuning through expected utility maximization (see text).

Examples of appropriate utility models for hearing aid patients include Q3 (Kates and Arehart, 2004), PHAQM (designed by John Beerends, TNO), PEMO-AQ (designed by Kollmeier's Oldenburg group, see also contribution by Bondy *et al.* (2007) at this conference) or alternatively the Speech Intelligibility Index (ANSI, 1997) if we choose that our utility should reflect intelligibility rates. Since $y=H(x,\theta)$, the utility model will implicitly depend on x and θ, so we will sometimes write $U(x,\theta)$ to emphasize this dependency. We define the *expected utility* (EU) by averaging the utility over a data set $X=\{x_1,\ldots,x_K\}$ of relevant input signals as follows,

$$\mathrm{EU}(\theta) = \frac{1}{K}\sum_{k=1}^{K} U(x_k,\theta) \qquad \text{(Eq. 2)}$$

The expected utility is a function of the HA tuning parameters θ. We will want to select HA tuning parameter values θ^* that maximize the expected utility. Mathematically, this idea can be expressed as

$$\theta^* = \arg\max_{\theta} \mathrm{EU}(\theta) \qquad \text{(Eq. 3)}$$

In principle, this idea can simply be executed by evaluating eq. 2 for a set of candidate values $\Theta=\{\theta_1,\ldots, \theta_J\}$ and choosing the value $\theta \in \Theta$ that gives maximum expected utility. While this plan for finding optimal HA tuning parameters is simple enough (although computationally heavy), there are a number of remaining problems with this approach. In particular,

1. What utility model should we use? In general, perceptual evaluations are different for each patient.
2. How do we make use of the measured patient data $\{D,a\}$?
3. How do we deal with inconsistent patient data and other uncertainties?

In the next sections we extend the basic HA tuning system (eqs. 2 and 3) to address these issues.

MODELS FOR PATIENT RESPONSES

A first observation is that perceptual preferences can be different for each patient. In order to model individual satisfaction rates we will use a parameterized utility model $U(y,\omega)$, where we allow the *utility parameters* ω to be adapted to individual patients.

The Fitting Model

While each patient will indeed be different from every other patient in some way, many patients will still perform similarly to others, in particular to those patients with similar auditory profiles. The assumption that *similar auditory profiles imply similar perceptual responses* can be captured by a *fitting model*

$$\omega = F(a,\psi) \qquad \text{(Eq. 4)}$$

where ψ are the parameters of the fitting model (to be discussed later), see Figure 2. A fitting model F takes as input an auditory profile a and provides as output the typical utility model for patients with similar profiles. The fitting model selects a typical utility model by setting the parameters ω of the utility model.

It may seem like a lot of unnecessary machinery to introduce yet another model F when we already have a utility model U. Still, we think that F is a very essential model that links behavioral data from the entire patient community to any specific patient. Thus, a properly tuned fitting model F may substantially reduce the amount of time needed to spend with any specific patient in order to get a well-tuned hearing aid algorithm.

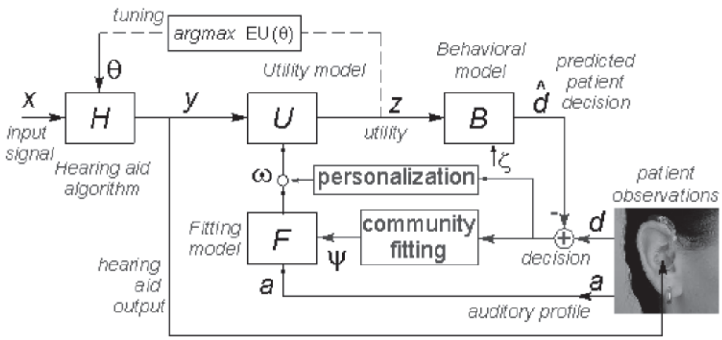

Fig. 2: Flow diagram for fitting and tuning (see text).

The Behavioral Model

Note that perceived utilities z are not the same as patient decisions during listening tests. A patient's perceived utility z is an internal representation and hence not observable. However, we can elicit utility-dependent decisions d from the patient through listening experiments and subsequently use these decisions to update our knowledge about the utility and fitting models. Maps from utilities to decisions are called *behavioral* (or *choice* or *decision*) models in the literature (Train, 2003), see Figure 2. In our framework, the behavioral model

$$\hat{d} = B(z, \zeta) \qquad \text{(Eq. 5)}$$

takes as input a set of utilities and delivers predicted patient decisions $\hat{d}$ as output. The particular choice for our behavioral model depends on the experimental protocol in the listening test. For instance, consider a paired comparison test where two signals y_1 and y_2 are presented to a patient who is asked to select the best signal. 'Best' here refers to the specific perceptual criterion that we wish to model. For instance, we may ask the patient to pick the signal with best 'overall quality', or with 'least amount of noise', etc. We can derive the behavioral model for patient decisions under the assumption that, when faced with a set of possible choices, a patient chooses the one with maximum perceived utility, i.e. a patient chooses y_1 if $z_1 > z_2$. Assume that the patient's decision is stored in a variable through the following assignment:

$$d = \begin{cases} 1 & \text{if } z_1 > z_2 \\ 0 & \text{else} \end{cases} \qquad \text{(Eq. 6)}$$

Under some mild mathematical conditions for the internal noise term ε_z, we can derive the following behavioral model for the probability that d=1,

$$\text{Prob}(d = 1) = \frac{1}{1 + \exp(-(z_1 - z_2))} \qquad \text{(Eq. 7)}$$

In the psychophysical community, this is called the Bradley-Terry model (Bradley and

Terry, 1952). For alternative experimental protocols, corresponding behavioral models can be derived.

COMMUNAL AND PERSONAL FITTING

Up to this point, we have introduced a few models that capture the perceptual responses of patients. Before we discuss how we train these models from measured patient data $\{D,a\}$, we describe the data set in more detail.

As discussed, we let the variable a hold a patient's auditory profile including his audiogram and SNR loss, but possibly also including variables such as gender, age or other demographic data. Basically, the variable a relates to all measurements on a patient excluding (behavioral) measurements obtained from listening experiments. A *listening experiment* (also known as a *trial* or *test*) consists of a set of *listening events* $\{e,d\}$, where e holds the inputs to a single listening event (e.g., a pair of audio samples), and d refers to the patient's response (decision) for the event (e.g., "first sample is preferred over second sample"). We will write D to denote the collective data obtained from a listening experiment. Consider a specific patient i. In general, there are two sources of behavioral data that we can use for learning about patient i's utility function.

(1) We can perform listening experiments on patient i (with auditory profile $a=a$) and collect the results in variable D_1.

(2) We can also use listening results from *all* other patients with similar auditory profiles and collect the results in the *community data* variable D_c.

Summarizing, there is a data set that contains data from a specific patient i and there is also a data set $\{D_c,a_i\}$, which refers to the collection of all other data sets. We also use the notation $D \int \{D_c,D_i\}$ to refer to all data from listening experiments and $a \int \{a_c,a_i\}$ for all available auditory profiles.

Now assume that we have collected a large community data set $\{D_c,a_c\}$ from listening experiments with corresponding auditory profiles. We can also run all audio samples from these listening experiments through our utility and behavioral model in order to produce a set of model-predicted patient responses $\hat{D}_c$. Next, we use a machine learning procedure to adapt the fitting model parameters to values $\psi=\psi_c$ such that the model predictions $\hat{D}_c$ match the actual patient responses D_c as closely as possible (see the box 'communal fitting' in Figure 2). The details of this learning procedure are beyond the scope of this paper. We call this process *communal fitting* because our fitting model F has now absorbed all relevant information from the community database $\{D_c,a_c\}$ into its coefficient values ψ_c.

Let us now assume that a trained fitting model has been installed in an appropriate fitting software product at a HA dispensing facility. Consider a patient   who visits a dispenser's office and is advised to accept a specific HA with algorithm $H(x,\theta)$. His auditory profile a_i is measured and used to select utility model parameters $\omega_i=a_i,\psi_c)$ that are typical for all other patients with similar auditory profiles. Next, for a relevant data set of audio samples $X=\{x_i,...,x_K\}$, we can compute the expected utility by eq. 2 and find optimal HA tuning parameters θ^* by executing eq. 3. At this point, we have used the data set $\{D_c, a_c, a_i\}$ to find the optimal tuning parameter values θ^* and effectively we have just designed the evidence-based HA algorithm $H(x,\theta^*)=H(x,D_c,a_c,a_i)$. Note that we have not yet performed any listening experiments with patient himself. Of course, we may decide to fine-tune the utility model to this particular patient through an individual listening test. In that case, for each audio sample, we adapt the utility model parameters ω further such that model predictions ($\hat{d}$) match actual responses (d) from patient i as closely as possible. Since this learning process is meant to individualize the utility model, we call this process *personal fitting* (or *personalization*). After personalization and expected utility maximization, we have effectively designed the evidence-based HA algorithm $H(x,\theta^*)=H(x,D_c,a_c,D_i,a_i)$. Note that in principle the procedure is automated and needs no guesswork. All we need are a HA algorithm $H(x,\theta^*)$ and a database of patient measurements $\{D,a\}$. Of course, we recognize that in commercial practice the proposed system would need to be integrated into an existing working practice in such a way that both dispenser and patients derive benefits. In Figure 3 we show a dependency graph of the variables in our system.

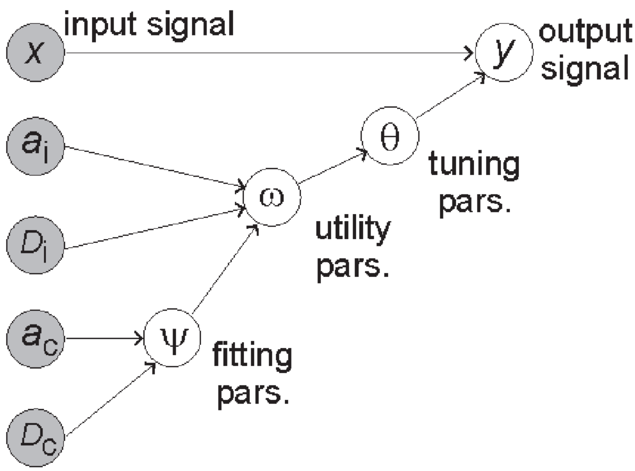

Fig. 3: Bayesian network representation of the evidence-based hearing aid algorithm $p(y \mid x,D,a)$ with listening test results $D=\{D_i,D_c\}$ and auditory profiles $a=\{a_i,a_c\}$. The shaded circles indicate observed variables $\{x,D_{,a}\}$; y is the HA output signal and the parameters $\{\theta,\omega,\psi\}$ are hidden (unobserved) variables, whose values are determined as intermediate results from all patient data.

DEALING WITH UNCERTAINTIES

An often-heard criticism of the machine learning approach is that “it won’t work because patient responses are inconsistent”. Indeed patient responses will often be inconsistent (e.g. different responses to the same listening event) which leads to uncertainty about the appropriate utility model and consequently uncertainty about the optimal tuning parameter values θ^*. In fact, there are more sources of uncertainty. For instance, when we compute the expected utility in eq. 2, how do we select a 'relevant' data set X of audio samples. Each choice of K samples consists of a (small) finite subset of the acoustic world and will be a biased representation leading to uncertainty about the 'true' expected utility.

The proper treatment of such uncertainties is to model them through probability distributions. Therefore, our entire system has been developed in a Bayesian probability theory context (De Vries *et al.*, 2006). In our probabilistic framework, we can actually measure and model the uncertainties and take them into account when deciding about optimal tuning values.

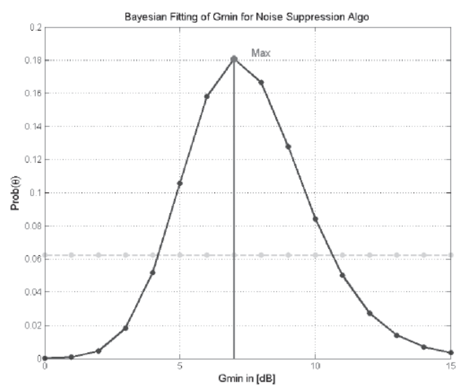

Fig. 4: PDF for preferred tuning parameter before (dashed) and after (solid) listening test.

In Figure 4, we show an example of the consequences of taking a probability theory approach to the problem. We took a noise suppression algorithm with one tuning parameter G_{min}. On the vertical axis, we plot the probability distribution function (pdf) for preferred values for G_{min} with respect to the expected utility. The utility parameters were initialized such that, before any listening event, all candidate values for G_{min} were equally likely to be preferred by the patient (see the dashed curve). Next, we conducted 70 paired comparison events for this algorithm. After training the utility model parameters ω to absorb the information from the listening test, the pdf for preferred values for G_{min} is peaked around 7 dB (see the solidly curved line in Figure 4). The width of the distribution indicates that the measured data implied some uncertainty about this optimal value G^*_{min}=7 dB. The data says that, e.g., G^*_{min}=8 dB would probably also be a good choice, but G^*_{min}=14 dB is unlikely to be a preferred choice. There is a lot more to be said about the probabilistic approach, but we summarize with the observation that the potential for modelling the inevitable uncertainties in the data leads to more informed decisions.

DISCUSSION

We have discussed a method for designing evidence-based HA algorithms, i.e. HA algorithms that take only observable-in-nature variables as inputs (and nothing else). Fundamentally, our approach takes as inputs a `regular' HA algorithm plus patient measurements , and produces as output an evidence-based algorithm . Accordingly, this method can be interpreted as a closed-loop fitting-engineering method because fitting results are produced as an intermediate result. The theory deals properly with data uncertainties through a Bayesian probability formalism.

As a final observation, note that our fitting model breaks with the current literature on hearing aid fitting models. In the hearing aid literature (and practice), it is common to develop direct maps from auditory profiles to HA algorithm tuning parameters . For instance, fitting rules like NAL and DSL map an audiogram directly to compression ratios, (e.g. Dillon, 2001 ch.9, for a review). While useful in practice, such rules suffer from a few serious drawbacks. For instance, the perceived compression depends on many algorithm tuning parameters including compression ratios, attack and release time constants as well as the presence of other modules in the complete HA algorithm. In principle, a new fitting rule should therefore be designed for each new HA algorithm. In our fitting approach, we develop a fitting map from auditory profiles to utility parameters . Thus, the fitting map relates to perceptual characteristics of patients rather than to hearing aid algorithms. This is proper: patient measurements say something about that patient (and in principle nothing about a HA algorithm), and should be used to update a model for that patient. Moreover, the perceptual properties of patients are more stable than HA algorithms. We do not need to develop a new fitting rule for each HA algorithm. The same fitting rule and utility model can be used to tune as many HA algorithms as we wish. Our system also leaves open the possibility to optimize several dependent HA algorithm parameters simultaneously (e.g., compression ratios and attack times).

REFERENCES

American National Standard Institute (ANSI-S3.5) (**1997**). "Methods for Calculation of the Speech Intelligibility Index."

Bondy, J., Coughlin, M., Whitmer, B., and Dittberner, A. (**2007**). "Assessing sound quality of feedback algorithms with auditory models," ISAAR 2007 conference, Helsingor, Denmark.

Bradley, R. A., and Terry, M. E. (**1952**). "Rank analysis of incomplete block designs. I. The method of paired comparisons," Biometrika, **39**, 324–345.

De Vries, B., Heskes,T. M., and Dijkstra, T. M. H. (**2006**). "Bayesian Incremental Utility Elicitation with Application to Hearing Aids Personalization," Valencia/ISBA 8th World Meeting on Bayesian Statistics , Benidorm, Spain.

Dillon, H. (**2001**). "Hearing aids," Boomerang Press, Australia. ISBN 1-58890-052-5.

Kates, J. M., and Arehart, K. H. (**2004**). "A metric for evaluating speech intelligibility and quality in hearing aids," J. of the Acoust. Soc. of America, Volume **116**, Issue 4, pp. 2536-2537.

Train, K. (**2003**). "Discrete choice models with simulation," Cambridge University Press.