The Bayesian Approach to Hearing Aid Fitting:
An Example with Common Sense Reasoning

Tjeerd M.H. Dijkstra, Alexander Ypma, Bert de Vries and Jos R.G.M. Leenen
GN ReSound A/S, Horsten 1, 5612 AX Eindhoven, The Netherlands
{tdijkstra,aypma,bdevries,jleenen}@gnresound.com

Abstract
This white paper discusses in a tutorial way the Bayesian probability theory approach to the problem of hearing aid fitting. Currently more an art than a science, we believe that probability theory will come to assist the dispenser in future generations of fitting software. We will show that probability theory is consistent with common sense reasoning, a feature that is not shared by alternative mathematical frameworks for intelligent reasoning. While probability theory gets to the same answers as a consistently reasoning human expert, it can deal with larger problems than a human expert. Since human expertise cannot be replaced by a mathematical system, we expect that mathematical reasoning systems will serve as an assistant to the dispenser in difficult fitting tasks.

1 Introduction

The Bayesian approach is gaining popularity in statistics, signal processing and artificial intelligence. For instance, in the design and analysis of clinical trials, use of Bayesian statistics is becoming accepted [FDA, 2006]. In signal processing, hearing aid algorithms are starting to make use of techniques from Bayesian machine learning, while in modern fitting software, more and more artificial intelligence techniques are embedded. These three areas of research are instrumental for improvements in hearing health care. Thus, we expect the hearing aid professional to be aided in increasing fashion with the Bayesian approach.

This paper exemplifies the Bayesian approach with a common problem in hearing aid fitting, namely how to set the gain in each frequency band. Starting from the scenario where a user returns to the dispenser with complaints about the hearing aid, we show how common sense reasoning can explain the issues. Subsequently, we imagine how a computer could reach the same explanation. We point out that this type of reasoning cannot be deductive based on classical logic but instead is inductive, based on probability theory, in particular Bayes rule (see Box 1). Scaling up this simple scenario to a state-of-the-art hearing aid algorithm with hundreds of tunable parameters, the problems cannot be solved by common sense human reasoning alone. Solutions are found in Bayesian machine learning which is designed to handle large quantities of uncertain data, using the same rules as common-sense human reasoning.

Box 1: who was Bayes and what is Bayes rule?

Rev. Thomas Bayes (1702-1761) was a minister in TunBridge Wells, England. Posthumously, his friend Richard Price published a paper entitled "An Essay Towards Solving a Problem in the Doctrine of Chances" that contains a version of what is nowadays called Bayes rule. Most students of statistics would not readily recognize this version as Bayes rule. That honor goes to the 1774 paper by the Frenchmen Pierre-Simon Laplace (1749-1827).

We will describe Bayes rule in the context of parameter estimation. Assume that an algorithm parameter Θ can be set to any of K values, namely θ_1, θ_2, …, or θ_K. We write

$$P(\Theta = \theta_k)$$

for the *plausibility* (probability) that θ_k is the preferred value for parameter Θ. For example, in the context of a noise reduction algorithm, $P(\Theta = \theta_k)$ could denote the plausibility of three possible settings, a small ($\Theta = \theta_1$), medium ($\Theta = \theta_2$) and large amount ($\Theta = \theta_3$) of noise reduction being preferred by a hearing aid user. Further assume that in general about half of the hearing aid users prefer the medium setting and that the small and large settings are preferred by a quarter of the users. Making that assumption quantitative, we assign $P(\Theta = \theta_1) = 0.5$ and $P(\Theta = \theta_2) = P(\Theta = \theta_3) = 0.25$. In determining the best setting for noise reduction, a dispenser performs listening tests. The outcome of a listening test is quantified by:

$$P(D = d_m \mid \Theta = \theta_k)$$

which denotes the plausibility of outcome $D = d_m$ given that the user prefers $\Theta = \theta_k$ (the vertical bar | is read as "given that"). For example if we assume that the three possible responses are "like" ($D = d_1$), "don't like" ($D = d_2$) and "don't know" ($D = d_3$), then $P(D = d_1 \mid \Theta = \theta_2)$ denotes the plausibility of response "like" given that this user prefers the medium noise reduction setting. What we would like to calculate is the plausibility that users prefers setting $\Theta = \theta_k$ using the observed data from the listening tests. This plausibility is denoted by $P(\Theta = \theta_k \mid D = d_m)$ and is calculated by **Bayes rule**:

$$P(\Theta = \theta_k \mid D = d_m) = \frac{P(D = d_m \mid \Theta = \theta_k)}{P(D = d_m)} \times P(\Theta = \theta_k)$$

The plausibility $P(D = d_m)$ is a normalization factor and is beyond the scope of this paper. Bayes rule allows the calculation of plausibilities one wants to know (on the left) from plausibilites that are known (on the right). Put another way, one has learned a new plausibility from the data. More information can be found in the wikipedia articles (http://www.wikipedia.org) on Bayes and Laplace.

2 A Hearing Aid Fitting Problem

A woman (Shirley) has just been fitted with a new hearing aid and has dinner that evening in a fine restaurant. Her dinner is somewhat disturbed by the clatter of her cutlery (knife and fork) on her plate. The sounds stand out more than what she is used to and it annoys her. A while later during the dinner, a live music band moves around the restaurant. The low resonating sounds of the tuba are uncomfortable. The next day she goes back to the dispenser and tells her story. The dispenser is about to make some adaptations to her hearing aid parameter settings. What reasoning process will he follow in order to draw conclusions?

2.1 Human reasoning and probability theory

Before continuing the Shirley example, we provide background on human reasoning, its relation to classical artificial intelligence (AI) and probability theory. In particular, we show that the rules of common-sense reasoning differ from deductive reasoning as used in classical AI. Starting with the latter, what are the rules of logical reasoning? Since Aristotle (4th century BC) it is known that all true statements can be derived using just two rules:

Rules of Deductive Logic

Rule 1 (*Premise*) A is TRUE, then B is TRUE
(*Fact*) A is TRUE
(*Conclusion*) B is TRUE

Rule 2 (*Premise*) if A is TRUE, then B is TRUE
(*Fact*) B is FALSE
(*Conclusion*) A is FALSE

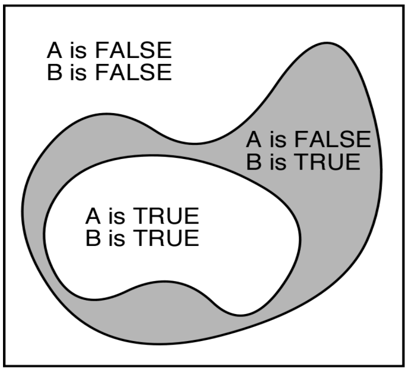

Figure 1. truth diagram for statements A and B

We should read these rules as follows. The statement "if A is TRUE, then B is TRUE" is the premise, the knowledge base. Suppose that a new fact becomes available, namely that A is TRUE. Then, according to Rule 1, it follows that B is TRUE. This kind of reasoning is called *deductive reasoning* and it forms the core of classical AI. Rule 2 says that, if we know that B is FALSE (instead of knowing that A is TRUE), it follows that A must be FALSE as well. No ifs and buts here, this is hard logic with binary outcomes: all statements are either TRUE or FALSE. Intuition for these deductive rules can be gained from the figure on the right-hand side. Since the area where B is TRUE encompasses the area where A is TRUE, it must be the case that B is TRUE whenever A is TRUE, which is essentially the same deduction as rule 1. Insight in rule 2 can be gleaned by noticing that the area where A is FALSE encompasses the area where B is FALSE.

We would love to use deductive reasoning for solving real-world problems. Unfortunately, many problems cannot be solved through deductive reasoning. Often we have the premise "if A is TRUE, then B is TRUE", combined with the observation "B is TRUE" or "A is FALSE". For instance, in the Shirley example we have the general rule "if the high-band gain is set too high then high-frequency noise will sound too loud" combined with the observation "high-frequency noise sounds too loud". Now what? There is no rule in deductive reasoning that we can use. We need to reason backwards from observations to possible causes, which is the problem of *induction*. In inductive logic we add two more rules:

Extra Rules of Inductive Logic

Rule 3 (*Premise*) if A is TRUE, then B is TRUE
(*Fact*) B is TRUE
(*Conclusion*) A becomes more plausible

Rule 4 (*Premise*) if A is TRUE, then B is TRUE
(*Fact*) A is FALSE
(*Conclusion*) B becomes less plausible

These rules do not allow one to reach conclusions with certainty like the rules of deductive logic. Instead, these rules express a *degree of plausibility* through increasing or decreasing the plausibility of statements. In daily life we often use these rules. For example, with premise: if it rains, is it cloudy. Fact: it is cloudy in New York. Hence, it is more plausible that it rains in New York. In daily parlance, these rules are called *common sense*. They can be made quantitative: we can associate a numerical value between 0 and 1 with the degree of plausibility of a statement. The truth value FALSE corresponds to the number 0 and the truth value TRUE corresponds to the number 1. Assigning numerical values to the degree of plausibility raises the question as to how the degrees of plausibility should be combined. Cox [1946] has shown that *Probability*

Theory is the only consistent mathematical system for combining all four rules of reasoning described above. This latter statement implies that any other mathematical system for doing intelligent reasoning is either consistent with probability theory (in which case it is redundant) or it is inconsistent (in which case it can lead to strange answers).

Probability theory is an abstract mathematical formalism that defines the notion of "probabilities" and rules for combining them. In brief, probabilities are quantities between 0 and 1 that are combined with the product and sum rules (which are outside the scope of this paper). A mathematical probability P(A) (i.e. a number between 0 and 1) can be interpreted as the *degree of plausibility* that statement A is TRUE. The rules from probability theory for combining degrees of plausibility then form a calculus of plausible reasoning. Bayes rule is a direct consequence of Probability Theory and hence provably a correct method for reasoning about degrees of plausibility. For a detailed discussion of these arguments see Jaynes [2003] and the articles on "deductive reasoning" and "inductive reasoning" on Wikipedia.

2.2 Human reasoning about a hearing aid fitting problem

We now show that the reasoning process that the dispenser follows cannot be based on deductive logic. He will use his common sense and we will show that this coincides with rule 3 and hence with inductive reasoning. We again consider Shirley, who has just been fitted with a new hearing aid. For the sake of simplicity we consider a hearing aid with just three tunable parameters, the gains in the low and high frequency band and the maximum output level (MPO), see Figure 2.

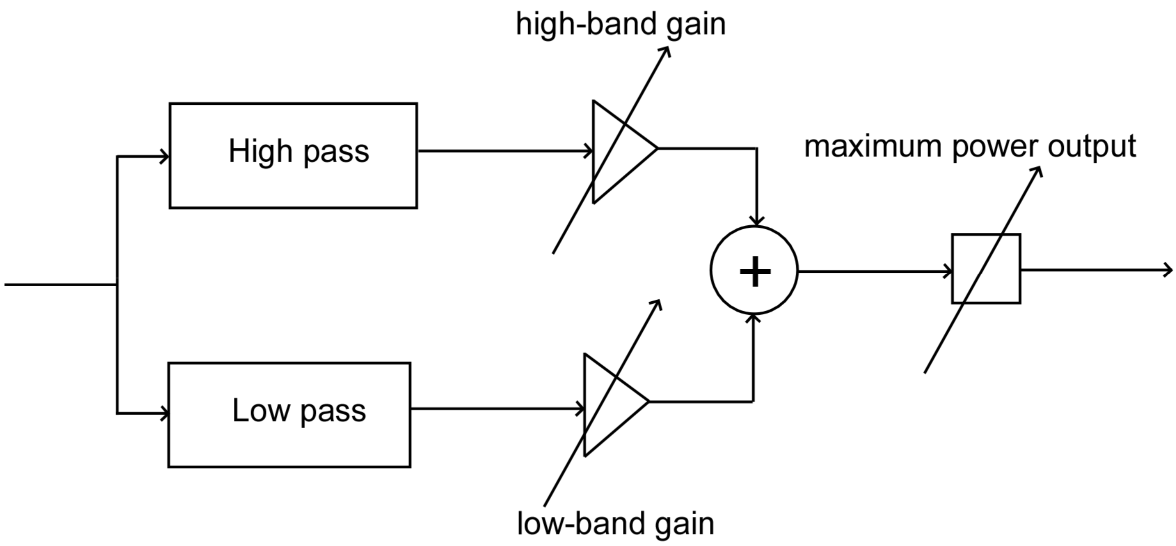

Figure 2: Schematic of Shirley's hearing aid algorithm with three parameters

Recapping the scenario, Shirley visited a restaurant and she experienced two problems with her hearing aid. First, she noticed that the clatter of her knife and fork sounds much more annoying than it used to. Second, there was a live band playing in the restaurant and she found the sound from the tuba very tiresome. The next day she returns to the dispenser and complains about her restaurant experience. The dispenser reasons as follows: Shirley has trouble both with the high (knife clatter) and the low frequencies (tuba). He has three possible hypotheses that are consistent with these observations:

(1) both the high-band gain and the low-band gain were set too high, or

(2) the MPO was set too high, or
(3) all three are too high.

Let us assume that the dispenser reasons that it is more plausible that one parameter is set to a wrong value (the MPO setting) rather than two parameters (both low-band and high-band gains) or three parameters (low-band and high-band gains and MPO). This is a *common sense* conclusion to a simple problem. Let us try to reason about this problem using deductive logic. As a first step, we assign labels to statements and start reasoning.

Possible cause LH = low-band gain is too high
Possible cause MH = maximum power output is too high
Possible cause HH = the high-band gain is too high
Observation TL = tuba sounds too loud
Observation UL = cutlery noise too loud

The expert knowledge of the dispenser can be represented by a set of rules:

Knowledge rule 1 = if LH is TRUE, then TL is TRUE
Knowledge rule 2 = if MH is TRUE, then TL and UL are both TRUE
Knowledge rule 3 = if HH is TRUE, then UL is TRUE

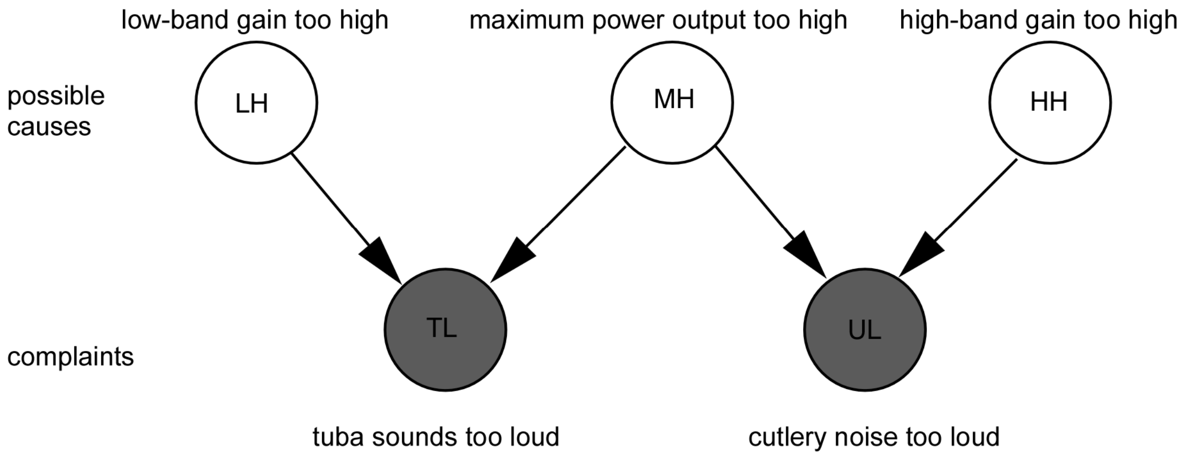

Figure 3: Bayes network for reasoning about Shirley's hearing aid problems. Open circles denote variables that are not observed and gray circles variables that are observed.

Figure 3 shows these statements and rules graphically in a so-called *Bayes Network*. The observations about the loud tuba and the clattering cutlery and are interpreted by the dispenser as evidence for observations TL and UL. Can he now deduct that LH is TRUE, or HH, or both LH and HH? No, he cannot: given that both low and high frequency sounds are annoying, he cannot conclude anything about the gain settings using deductive logic. The rules of deductive logic (rules 1 and 2) do not apply here. Even in this simple problem he must apply inductive logic. Since TL is observed, he concludes from inductive rule 3 that LH (low-band gain too high) becomes more plausible. UL is also observed hence he concludes that HH (high-band gain too high) is also more plausible. Since TL and UL are both observed, he concludes that it is more plausible that the MPO was set too high.

So all possible causes are more plausible. But which one is most plausible? Intuitively, it seems that the explanation based on the level of the maximum power output is the most plausible, as it is supported by two observations, whereas the other explanations are only supported by one observation and seem more coincidental. As discussed, we need Bayes rule to draw quantitative conclusions about the relative plausibility of the possible causes. As calculated in the next section, Bayes rule shows that the most plausible cause is indeed that the maximum power output is set too high. But first, Box 2 describes the relation between Bayesian and conventional (frequentist) statistics, and how these concepts from each can be used to build models for real-world problems.

Box 2: Three kinds of statistics

Most people think there is just one kind of statistics, the kind you loved to hate in college. However, one can discriminate at least three different approaches to statistics in use in modern practice. Why are there different kinds of statistics at all? A simple answer to this question is that statistics is about the relationship between the mathematical formalism of probability theory and the real world. Since one does not know the real world (only the measured data) one has to make assumptions about it … different assumptions lead to different kinds of statistics.

Frequentist statistics has been the dominant type in the 20th century and is the kind we learned in college. It goes back to Neyman and Pearson who advocated its use in the 1930-ies. It is called "frequentist" since it interprets the outcome of an experiment as one of an infinite number of potential experiments. The familiar *p*-value is then the frequency with which one can expect an outcome in the set of potential repeated experiments. Its advantages are its computational simplicity, an important bonus in the pre-computer era and its objectivity. Its disadvantage is the lack of flexibility and its peculiar interpretation: why does one need to imagine an infinite set of potential experiments just to interpret a single experiment?

Bayesian statistics was the dominant type in the 19th century and has started to gain popularity in the 21st century. The advantages are highlighted in the main text. One disadvantage is the computational burden although this is hardly an issue with modern computing power. A second disadvantage could be that it is more work since one needs to specify priors (the $P(\Theta = \theta_k)$ in box 1) for all relevant parameters. Alternatively, one could view this as an advantage since it forces the modeler to be explicit about her assumptions. To confuse matters, even within the Bayesian community, there are several distinct viewpoints on how to interpret the concept of probability. In this paper, we adhere to the interpretation of probability as a state-of-knowledge. This is the original viewpoint of Bayes and Laplace and further developed by Jeffreys (1939), Cox (1946) and Jaynes (2003).

Machine learning is a modern branch of applied statistics and uses concepts from both frequentist and Bayesian statistics. Machine learning focuses on larger problems with practical applications in mind and mixes in methods from other fields like Artificial Intelligence and Numerical Analysis. Bayesian machine learning is a subset of machine learning that uses Bayes rule for learning model parameters. Models built with Bayesian machine learning combine prior (expert) knowledge about a problem with measured data. We advocate the use of Bayesian machine learning as a framework for building consistent reasoning systems in hearing aids.

2.3 Bayesian reasoning about a hearing aid fitting problem

In this section we provide a quantitative analysis of Shirley's problem. Necessarily, this section is more technical than the previous ones. Thus, a reader who is content with the gist of the paper can skip section 2.3. However, we provide a hands-on implementation of the Shirley example as a Bayes network using the freeware package GeNIe, allowing a reader to experiment with different choices of the parameters.

For full specification of the Bayes network as depicted in figure 3, we need 11 numbers: three for the prior probabilities on each of the possible causes and eight for the complaints (four for each complaint). The three priors denote how certain the dispenser is that he made the correct fit. The dispenser generally fits a hearing aid correctly thus, for the purpose of exposition, we set the probabilities for all causes to a small number, say 0.1. In formula:

$$P(LH = TRUE) = 0.1$$
$$P(MH = TRUE) = 0.1$$
$$P(HH = TRUE) = 0.1$$

Next we specify the degree of plausibility that a complaint occurs for all possibilities of the causes. Focusing on the complaint that the tuba sounds too loud, we reason that when both the low-band gain is too high and the maximum power output is too high, it is quite certain that the tuba will sound too loud, say 0.9. In formula:

$$P(TL = TRUE \mid LH = TRUE \text{ and } MH = TRUE) = 0.9$$

Similarly, when just either of the possible causes is true we assume it is a bit less likely that the tuba sounds too loud, say 0.8. In formula:

$$P(TL = TRUE \mid LH = TRUE \text{ and } MH = FALSE) = 0.8$$
$$P(TL = TRUE \mid LH = FALSE \text{ and } MH = TRUE) = 0.8$$

Lastly, when neither of the two possible causes is TRUE it is rather unlikely that the tuba sounds too loud, say 0.1. In formula:

$$P(TL = TRUE \mid LH = FALSE \text{ and } MH = FALSE) = 0.1$$

We assume the network is symmetric, and thus the values are the same when the cutlery noise is too loud. In formula:

$$P(UL = TRUE \mid LH = TRUE \text{ and } MH = TRUE) = 0.9$$
$$P(UL = TRUE \mid LH = TRUE \text{ and } MH = FALSE) = 0.8$$
$$P(UL = TRUE \mid LH = FALSE \text{ and } MH = TRUE) = 0.8$$
$$P(UL = TRUE \mid LH = FALSE \text{ and } MH = FALSE) = 0.1$$

We have now fully specified our knowledge and are ready to reap the benefits. Although it is straightforward to calculate the plausibility of the possible causes by Bayes rule, it more convenient to use the freeware package GeNIe. GeNIe can be downloaded for free from http://genie.sis.pitt.edu/ after a simple registration process. We have implemented the Shirley example as a Bayes network that can be visualized in the GeNIe environment, see Figure 4. Using GeNIe we find a plausibility of 0.716 for the possible cause that the maximum power output is too high and 0.213 for the alternative possibility that the low-band gain is too high. The possibility for the high-band gain being too high is also 0.213 since the network is symmetric by design. Thus, the level of the maximum power output comes out as the most plausible cause, as expected.

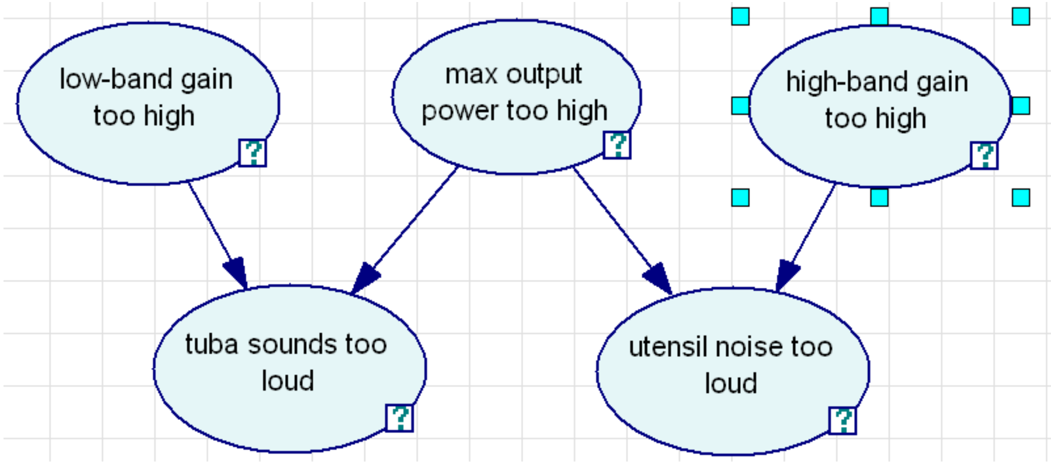

Figure 4: Screendump of the Shirley network in GeNIe

Given the relative complexity of the calculations, it seems natural to ask whether this overhead is necessary to reach such a straightforward conclusion. A first answer to this objection is that real problems have many more variables and cases, which limits the use of intuition. In contrast, the Bayesian calculations scale to much bigger problems with many variables. A second answer is that there is an important payoff in the Bayesian approach: one can update the knowledge base with the knowledge that is acquired while diagnosing hearing aid fitting problems. This is explained in more detail in the next section.

3 Learning of a Dispenser Knowledge Base as Bayesian Prior Updating

In the previous section we found that the Bayesian approach allows the dispenser to diagnose Shirley's problem correctly. One ingredient of the approach is the so-called prior that constitutes the dispenser's belief about the plausibility of incorrectly fitting a patient. The word "prior" refers to the notion that this is the dispenser's initial belief before having diagnosed Shirley. In the example we set these priors to 0.1 for each of the possible causes. We would now like to update the dispenser's prior beliefs since the dispenser has found another incorrect fit: Shirley's hearing aid needed adjustment. Obviously, the dispenser needs to increase his prior for the possible cause of setting the maximum output level too high. By how much should the dispenser increase his prior? Bayes comes to the rescue.

In fact, it is not so much Bayes as Laplace who came up with the answer, although it is based on Bayes rule explained in Box 1. We need one extra piece of information and that is the number of patients on which the dispenser's prior is based. Say that the dispenser has seen 18 patients before Shirley and that one of them returned with problems similar to Shirley. According to *Laplace's rule of succession*, the prior plausibility should be set to the number of misfits plus one (making a total of 2) divided by the total number of patients plus 2 (18+2 = 20 -- why the extra +1 and +2 are needed is outside the scope of this paper). Thus, the prior plausibility is 2/20=0.1 for the possible cause of MPO too high.

As Shirley is the nineteenth patient seen by the dispenser, we can use Laplace's rule of succession to update the prior. The number of misfits is now 2 and the number of patients 19 leading to an updated prior of 3/21 which equals about 0.143. Obviously, the number of previously seen patients influences how much the prior is updated by Shirley's case. If the dispenser had seen 998 patients previously, 99 of whom had returned and needed an adjustment of the maximum power output, then the prior would be updated from 100/1000 to 101/1001 which is a small change. One can view this updating of prior information as the learning of a knowledge base. The learned knowledge base contains information about the dispenser and his clients. As a consequence of the learning, if another person came along with identical complaints and an identical hearing aid, the fitting system would suggest that the MPO is too high with a plausibility larger than the 0.716 we calculated previously. Nothing in the mathematics of our example is specific to hearing instrument counseling and could just as well be applied in other areas (see Box 3).

Box 3: Other applications of Bayesian statistics

The Bayesian approach is gaining acceptance in clinical evaluations. In particular, the Food and Drug Administration (FDA) in the USA has just issued a guidance document [FDA, 2006] allowing the use of Bayesian statistics in the clinical evaluation of medical devices, which includes hearing aids. The guidance document lists the following advantages of Bayesian statistics over frequentist statistics:

Sample size reduction or augmentation. The Bayesian methodology may reduce the sample size FDA needs to reach a regulatory decision. You may achieve this reduction by using prior information and interim looks during the course of the trial. When results of a trial are unexpectedly good (or unexpectedly bad) at an interim look, you may be able to stop early and declare success (or failure). The Bayesian methodology can allow for augmentation of the sample in cases where more information helps FDA make a decision. This can happen if the observed variability of the sample is higher than that used to plan the trial.

Midcourse changes to the trial design. With appropriate planning, the Bayesian approach can also offer the flexibility of midcourse changes to a trial. Some possibilities include dropping an unfavorable treatment arm or modifications to the randomization scheme. Modifications to the randomization scheme are particularly relevant for an ethically sensitive device or when enrollment becomes problematic for a treatment arm. Bayesian methods can be especially flexible in allowing for changes in the treatment to control randomization ratio during the course of the trial.

Exact analysis. The Bayesian approach can sometimes be used to obtain an exact analysis when the corresponding frequentist analysis is only approximate or is too difficult to implement.

The Bayesian approach has led to various **commercial applications outside the hearing aid industry**. Examples can be found in diagnostics, like the Promedas system for medical diagnosis [Kappen, Wiegerinck and ter Braak, 2002] and the SKF BearingInspector system, a system for identifying bearing failures [van der Vorst and Schram, 2003]. Furthermore, Bayesian reasoning is at the basis of systems that have been used at NASA, e.g. the decision-theoretic Vista system to support Mission Control, [Horvitz and Barry, 1995] and several tools for automated data analysis and super-resolution imagery. Examples in the realm of computers are: Microsoft Lumière, the technology behind the user-requested services provided by the Office Assistant in MS Office and Outlook Mobile Manager, a system that evaluates incoming e-mail messages and decides which ones are important enough to forward to a pager, mobile phone or other e-mail address [The Economist, 2001]. Other applications that are often 'powered by' Bayesian Networks include speech recognition, robot vision, genetic linkage analysis, numerous tracking applications and turbo-coding (turbo-codes are part of third-generation mobile telephony protocols). Bayesian machine learning was listed as one of *10 emerging technologies that will change your world* in the February 2004 edition of MIT's Technology Review [Technology Review, 2004].

We have exemplified the Bayesian approach with a scenario from clinical practice of a dispenser, where a person returns with a complaint about the hearing instrument. We stress that the example in no way implies that the dispenser did anything incorrect. Fitting of hearing instruments is an art more than a science and professional care will always be required. What the Bayesian approach brings is a bit more science to the art.

4 User Preference Learning as Bayesian Prior Updating

As discussed in the introduction, the Bayesian approach could also be used in the signal processing going on inside a hearing instrument. To illustrate this facet, we discuss learning of the preference of the volume control: the learning volume control. Learning of a volume preference has been discussed before by [Dillon, Zakis, McDermott, Keidser, Dreschler and Convery, 2006], [de Vries, Ypma, Dijkstra and Heskes, 2006], [Ypma, de Vries and Geurts, 2006] and [Chalupper and Powers, 2006]. To simplify, we consider a learning volume control with only three levels +4 dB, 0 dB and -4 dB. We also imagine a sound classifier inside the hearing instrument that classifies the incoming sounds into discrete categories, for example, speech, music and other. Many modern hearings instruments have a sound classifier in their signal processing. Since the classifier is not the focus of this paper, we simply assume that one exists. The question is how to associate a level of the learning volume control (say +4 dB) with a sound class (say music).

Things would be simple if users always preferred a certain level for a certain sound class. In that case one could run a set of tests to establish user preference and be done. Unfortunately, sound classifiers make errors, users are not always consistent and may change their preferences over time. Consider the trainable hearing aid as proposed by [Dillon et al, 2006]: an instrument that uses feedback from the user to adapt its settings. Dillon et al. propose the concept without specifying how the instrument can learn from user input. The purpose of this section is to illustrate how the Bayesian approach could do that.

Following the Bayesian approach, we associate a plausibility with each setting of the learning volume control. For example, P(LVC = +4 dB | sound = music) denotes the plausibility that the user prefers an extra 4 dB of amplification in the situation that the sound is classified as music. Since the mathematics is identical for each sound class, we assume the class is given and leave it out of the notation. Thus for each sound class we have three numbers P(LVC = +4 dB), P(LVC = 0 dB) and P(LVC = -4 dB), expressing the plausibility that the user prefers the associated level of amplification.

We imagine that the user can express preference for a certain level of the added gain by pressing a button repeatedly or by turning a volume control wheel. While the Bayesian approach also applies to a volume control wheel, the principles are more easily explained for a volume control with discrete levels. Further, imagine that the user presses the button once expressing preference for the extra 4 dB of amplification. Clearly, we should increase the plausibility for P(LVC = +4 dB) and decrease the other two since all three should sum to 1. By how much should P(LVC = +4 dB) be increased? By now, the answer should have a familiar ring to it: Bayes theorem tells us by how much.

The problem is identical to the prior updating in the Shirley example. Thus, we need to keep track of how often the user preferred a certain level of the learning volume control and use Laplace's rule of succession: the plausibility is the number of button presses plus 1 divided by the total number of button presses plus three. If this is the first button press by the user, there have been zero presses for the other two plausibilities and we get:

P(LVC = +4 dB) = (1 + 1) / (1 + 3) = 0.50
P(LVC =   0 dB) = (0 + 1) / (1 + 3) = 0.25
P(LVC = -4 dB) = (0 + 1) / (1 + 3) = 0.25

Which level of amplification should the user hear? The answer to this question is outside the scope of the Bayesian approach per se and is the realm of decision theory. Two possibilities are the winner-take-all approach where the user hears 4 dB of amplification or the weighted average approach where the user hears 0.5*(+4) + 0.25*0 + 0.25*(-4) = 1 dB.

The avid reader might have noticed that we gave two versions of Laplace's rule of succession: in the previous example of updating the dispenser prior we added two to the denominator and here we add three. Which is correct? The answer is that both are correct, as the number added to the total number of cases in the denominator should be the number of plausibilities considered which was two in the Shirley example and is three here.

The reader might have another objection to this formalism. Why use the Bayesian approach with Laplace's rule of succession at all? It is much simpler to calculate the plausibility from the number of button presses for each level of the learning volume control divided by the total number of button presses:

P(LVC = +4 dB) = 1/1 = 1.0
P(LVC =   0 dB) = 0/1 = 0.0
P(LVC = -4 dB) = 0/1 = 0.0

Combined with the winner-take-all decision rule there is no difference between the simple approach and the Bayesian approach. Using the weighted average decision rule there is a difference but is that all the advantage. Luckily, Bayes has one more twist.

When sending a user off with a new hearing instrument, a dispenser would like to initialize the instrument to the best values. In the scenario above, it seems that best values are zero button presses for each of the three preferences. As a consequence, when the user has expressed his preference only a few times, large changes in the amplification are possible. One could say that this reflects the user's preference and be done. Alternatively, one could obtain more information to bear on the matter of initialization. Imagine that the hearing instrument company has run a field test with the instruments and found that users preferred the +4 db amplification setting with a mean P(LVC = +4 dB) of 0.6 and a standard deviation of 0.2. How do we use this information in initializing the learning volume control? Somehow, we have to convert to the 0.6 and 0.2 (learned from the field trial) to a number of button presses. Since the user has not actually pressed the button yet, these are imagined button presses, indicating the most likely preference of the user. In Bayesian parlance, these button presses are called *prior information*. A calculation beyond the scope of this paper gives 3.6 prior button presses for the +4 dB preference and 1.2 for the other two. Note that the button presses now are fractional numbers. Does that matter? No not really, these fractional prior button presses enter the calculation just as real button presses. What happens if a particular user prefers the 0 dB setting? Nothing other than that this user needs to indicate his preference for the 0 dB setting at least three times so that the prior for +5 dB (= 3.6) is overruled by the preference for 0 dB (1.2 + 3 = 4.2).

Summarizing, the Bayesian approach can be seen as a way to integrate information. Here we have seen how to integrate the information from one button press with information from previous presses and how to integrate information from a field trial with information obtained from an individual user.

5 The GN ReSound approach

Modern hearing aids have hundreds of parameters under the hood. A dispenser can no longer use only common sense and expert knowledge to cope with all the complexities and draw the optimal conclusions about parameter settings. GN ReSound is working on **a Bayesian Fitting Assistant** that suggests settings for the hearing aid algorithm parameters based on inputs such as listening tests, auditory profile measurements and expert knowledge, e.g. [Heskes and de Vries, 2005] and [de Vries, Ypma, Dijkstra and Heskes, 2006]. We envision this assistant to be an add-on to existing fitting software: a tool to aid the dispenser in optimizing the fitting and allowing the dispenser to better counsel his clients.

Using essentially the same mathematics, GN ReSound is also working **on a Bayesian approach to integrate feedback information from the user to steer algorithms**, e.g. [Ypma, de Vries and Geurts, 2006]. We envision that users can provide rich feedback about their preferences to a future generation of hearing aids, either through a remote control during use or through listening tests in the dispensers' office.

6 Conclusions

We have made a case for Bayesian probability theory as the foundation for solving hearing care problems in a quantitative matter. In 1812, Laplace made the following observation:

> "The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which oftentimes they are unable to account." [Laplace, 1812].

Indeed, we have demonstrated with a simple example –- the Shirley example— that Bayesian probability theory is consistent with common sense. The fitting of a hearing aid with many tunable parameters is too complex for common sense reasoning, but computers carry out the calculations and assist the dispenser. An important benefit from Bayesian probability theory is that the system can learn from data by updating the prior. We illustrated this in both the Shirley example by learning an updated prior as in the learning volume control example by learning an updated user preference. In the latter case we also showed how to integrate information from a field trial in the hearing aid. In summary, we proposed the plausibility from Bayesian probability theory as a currency to express knowledge and learning as the updating of these plausibilities. In the future, we anticipate learning like this to take place both in a fitting assistant as in the hearing instrument itself and provide a new tool for the dispenser.

References

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

de Vries, B., Ypma A. Dijkstra T. M. H., and Heskes T.M., *Bayesian Machine Learning for Personalization of Hearing Aid Algorithms*, Int. Hearing Aid Research Conference, Lake Tahoe (CA), August 2006.

Chalupper J. and Powers A.P., *Changing how gain is selected: the benefits of combining datalogging and a learning VC*, The Hearing Review, 13, 46-55, 2006.

Cox R.T., *Probability, Frequency, and Reasonable Expectation*, American Journal of Physics, 14, p 1-13, 1946.

Dillon K., Zakis J.A., McDermott H., Keidser G., Dreschler W. and Convery E., *The trainable hearing aid: What will it do for clients and clinicians?*, The Hearing Journal, 59, 30-36, 2006.

FDA (USA Food and Drug Administration) (2006), *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials - Draft Guidance for Industry and FDA Staff*, Docket number 2006D-0191.

Heskes T. M. and de Vries B., Incremental Utility Elicitation for Adaptive Personalization, Belgian-Dutch Conference on Artificial Intelligence (BNAIC), Brussels (Belgium), October 2005, pp. 127-134

Horvitz E. and Barry M., *Display of information for time-critical decision making*, In: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, pp. 296-305, 1995.

Jaynes E.T., *Probability theory: the logic of science*, Cambridge University Press, 2003. The first three chapters of this book are available on-line.

Jeffreys H., *Theory of probability*. Clarendon Press, 1939; later editions 1948, 1961, 1967, 1988.

Kappen H.J., Wiegerinck W. A. J. J. and ter Braak, E., Decision support for medical diagnosis, In: Dealing with the data flood – mining data, text and multimedia, pp. 111-121.

Laplace P.S., "Théorie analytique des probabilités", 1812.

Technology Review, *10 Emerging Technologies That Will Change Your World*, MIT, February 2004

The Economist (print edition), Son of paperclip, 22 March, 2001.

van der Vorst B. and Schram G., Decision-support system for bearing failure mode, Evolution – SKF Business and Technology Magazine, #1/03, 6 March 2003.

Ypma A., de Vries B. and Geurts J., Robust Volume Control Personalization from On-line Preference Feedback, IEEE Int. Workshop on Machine Learning for Signal Processing, Maynooth (Ireland), September 2006, pp. 441-446