A Gaussian process mixture prior for hearing loss modeling

Marco Cox[1] M.G.H.COX@TUE.NL
Bert de Vries[1,2] BDEVRIES@IEEE.ORG
[1]Dept. of Electrical Engineering, Eindhoven University of Technology, [2]GN Hearing, Eindhoven.

Keywords: hearing loss, probabilistic modeling, Bayesian machine learning.

1. Introduction

The most common way to quantify hearing loss is by means of the *hearing threshold*. This threshold corresponds to the lowest sound intensity that the person in question can still perceive, and it is a function of frequency. The typical process of measuring the hearing threshold is known as *pure-tone audiometry* (Yost, 1994), and it usually consists of incrementally estimating the threshold value at a set of standard frequencies ranging from 125 Hz to 8 kHz using a staircase "up 5 dB – down 10 dB" approach.

A recent line of work in the field of machine learning has focused on improving the efficiency of hearing loss estimation by taking a probabilistic modeling perspective (Gardner et al., 2015b; Song et al., 2015; Gardner et al., 2015a). This approach assumes that the hearing threshold of a person is drawn from some prior probability distribution. Under this assumption, the estimation problem reduces to a (Bayesian) inference task. Since the resulting posterior distribution describes both the estimated threshold and its uncertainty, it is possible to select the 'optimal' next test tone based on information-theoretic criteria. The so-called *active learning loop* (Cohn et al., 1996) of repeatedly selecting the best next experiment and updating the probabilistic estimate, significantly reduces the total number of required test tones (Gardner et al., 2015b).

The success of the probabilistic approach hinges on the selection of a suitable hearing loss model. Presently, the Gaussian process (GP) model is the best-performing model of the hearing threshold as a function of frequency (Gardner et al., 2015b). A GP can be viewed as a probability distribution over the space of real-valued functions (Rasmussen & Williams, 2006).

In this abstract we introduce a prior distribution for hearing thresholds learned from a large database containing the hearing thresholds, ages and genders of around 85,000 people. Almost all existing work is based on very simple and/or uninformative GP priors; simply selecting a suitable type of kernel that assumes the threshold curve to be smooth is already sufficient to yield a well working system. However, by fitting a slightly more complex model to a vast database of measured thresholds, we obtain a prior that is more informative and empirically justified.

2. Probabilistic hearing loss model

The hearing threshold is a (continuous) function of frequency, denoted by $t : \mathbb{R} \rightarrow \mathbb{R}$. The goal is to specify an appropriate prior distribution $p(t|a, g)$ conditioned on age $a \in \mathbb{N}$ and gender $g \in \{\text{female}, \text{male}\}$. We choose $p(t|a, g)$ to be a Gaussian process mixture model in which the mixing weights depend on age and gender:

$$p(t|a, g) = \sum_{k=1}^{K} \pi_k(a, g)\mathcal{GP}(t|\boldsymbol{\theta}_k). \qquad (1)$$

All K GPs have independent mean functions and kernels, parametrized by hyperparameter vectors $\{\boldsymbol{\theta}_k\}$. In our experiments we use third-order polynomial mean functions and the squared exponential kernel, which enforces a certain degree of smoothness on the threshold function, depending on its length-scale parameter. We do not fix mixing function $\pi(\cdot, \cdot)$ to a specific parametric form, but use a nearest neighbor regression model.

The main idea behind the choice for a mixture model is that it seems reasonable to assume that hearing thresholds can roughly be classified into several types. These types would correspond to different degrees of overall hearing loss severity, as well as hearing loss resulting from different causes, i.e. natural ageing versus extensive exposure to loud noises. The audiology literature indeed describes sets of "standard audiograms" to this end (Bisgaard et al., 2010).

3. Model fitting and evaluation

We fit the model parameters – GP hyperparameters $\boldsymbol{\theta}_1$ through $\boldsymbol{\theta}_K$ and mixing function $\pi(a, g)$ – to a database containing roughly 85k anonymized records from the Nordic countries. Each record contains the age and gender of the person in question, together with the hearing thresholds of both ears measured at (a subset of) the standard audiometric frequencies. The total set of 170k threshold measurement vectors is randomly split into a training set (80%) and a test set (20%) for performance evaluation.

The inference algorithm consists of two parts. Since all threshold measurement vectors correspond to a fixed set of frequencies, the GP mixture reduces to a mixture of multivariate Gaussians. Therefore, in the first part we fit a Gaussian mixture model to the training set using the expectation maximization algorithm (Moon, 1996). In the second part, we find the optimal GP hyperparameter values by minimizing the Kullback-Leibler divergence between the GP mixture and the multivariate Gaussian mixture using gradient descent.

Figure 2 visualizes the fitted prior conditioned on different ages. The means of the mixture components indicate that different components indeed capture different types of threshold curves. Moreover, conditioning the prior on age has a clearly visible impact. This impact is quantified in Figure 1, which shows the average log-likelihood of hearing thresholds in the test set. It also shows that the GP mixture priors outperform the empirically optimized single GP prior in terms of predictive accuracy.

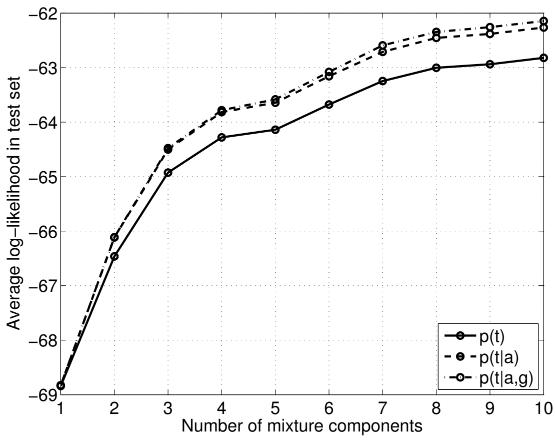
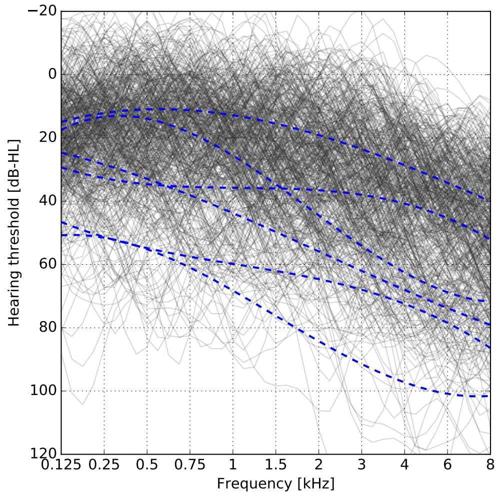
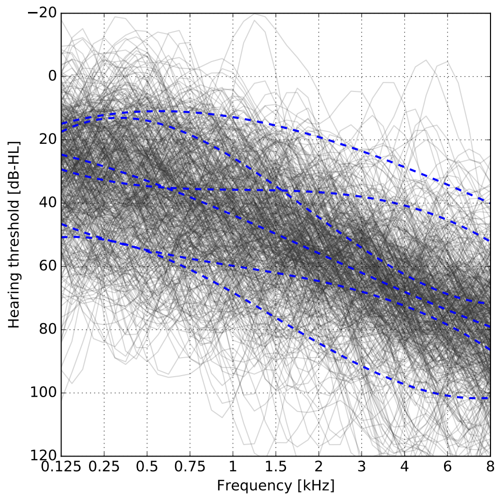


Figure 1. Predictive performance of the fitted priors on the test set. The one mixture component case corresponds to a standard GP prior with empirically optimized hyperparameters. Conditioning on age and/or gender consistently improves the predictive accuracy.



(a) $p(t|a = 40)$



(b) $p(t|a = 80)$

Figure 2. Visualization of the learned prior for $K = 6$ mixture components, conditioned on different ages. Blue dashed lines indicate the means of the mixture components. The gray lines are samples from the conditional priors. A value of 0 dB-HL corresponds to no hearing loss.

4. Conclusions

We obtained a prior for hearing loss by fitting a GP mixture model to a vast database. Evaluation on a test set shows that the mixture model outperforms the (empirically optimized) GP prior used in existing work (Gardner et al., 2015b), even without conditioning on age and gender. If age and gender are observed, the prior consistently becomes more informative. The benefit of adding more components to the mixture tapers off after about eight components.

References

Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in amplification*, *14*, 113–120.

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of artificial intelligence research*, *4*, 129–145.

Gardner, J., Malkomes, G., Garnett, R., Weinberger, K. Q., Barbour, D., & Cunningham, J. P. (2015a). Bayesian active model selection with an application to automated audiometry. *Advances in Neural Information Processing Systems* (pp. 2386–2394).

Gardner, J. R., Song, X., Weinberger, K. Q., Barbour, D. L., & Cunningham, J. P. (2015b). Psychophysical Detection Testing with Bayesian Active Learning. *UAI* (pp. 286–295).

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, *13*, 47–60.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* MIT Press.

Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q., & Barbour, D. L. (2015). Fast, continuous audiogram estimation using machine learning. *Ear and hearing*, *36*, e326.

Yost, W. A. (1994). *Fundamentals of hearing: An introduction (3rd ed.)*, vol. xiii. San Diego, CA, US: Academic Press.