

# Computational Models of Perceptual Learning Across Multiple Auditory Tasks: Modeling Daily Learning Limits as Memory Decay

**David Little (d-little@u.northwestern.edu)**

Department of Electrical Engineering and Computer Science  
Northwestern University, Evanston, IL 60208, USA

**Bryan Pardo (pardo@northwestern.edu)**

Department of Electrical Engineering and Computer Science  
Northwestern University, Evanston, IL 60208, USA

## Abstract

Humans have a remarkable ability to adapt their perceptual acuity to the task at hand, commonly referred to in the literature as perceptual learning. Understanding this ability at a computational level may have important implications across a wide variety of different psychological phenomena. There is evidence suggesting this ability plays an important role in speech comprehension, mathematics, and perceptual expertise, for instance. Computational models of perceptual learning have largely focused on hypothesizing how one or more mechanisms might explain the observed perceptual learning for a single task. Here we explore how a single model might explain the learning curves across two auditory perceptual learning tasks. Our results suggest that an ideal observer model with noisy input can predict learning when daily limits are not reached, and that daily limits on learning can be modeled by a decay of memory for trials observed on the current day of practice.

**Keywords:** perceptual learning; ideal observer; plasticity vs. stability; frequency discrimination; duration discrimination; temporal interval discrimination

## Introduction

Humans have a remarkable ability to adapt their perceptual acuity to the task at hand, commonly referred to in the literature as perceptual learning (Fahle and Poggio, 2002). Perceptual learning has been demonstrated in many different experiments. In vision for instance, there are studies of vernier hyper-acuity (Poggio et al., 1992), orientation discrimination, and spatial frequency discrimination (Fiorentini and Berardi, 1980). Examples in the auditory domain include results for frequency discrimination (Demany, 1985), and temporal interval discrimination (Wright et al., 1997). Perceptual learning is often characterized as being highly specific both to the task (Fiorentini and Berardi, 1980), and to the specific location or range within a dimension (Wright and Zhang, 2009; Poggio et al., 1992).

There is evidence that perceptual learning is important for a great variety of real world tasks humans face (Kellman and Garrigan, 2008). There is data suggesting that perceptual learning helps us during speech comprehension (Norris et al., 2003), that it can help children with dyslexia (Hayes et al., 2003) and that it has an important role to play in the comprehension of mathematical formulae (Kellman et al., 2008).

Computational models of perceptual learning have the potential to enable better predictions and to help us better understand human data. Past computational work studying percep-

tual learning has largely focused on how specific mechanisms might explain the particular properties of perceptual learning for a single task (e.g. Poggio et al., 1992; Petrov et al., 2005; Jacobs, 2009). Such studies focus on the question of how and/or where perceptual learning occurs within the human brain for a *single* perceptual task. Our goal here is to develop a model of *multiple* perceptual learning tasks. By looking across several tasks we can ultimately constrain our model by requiring that a single parameter explain qualitatively different results across several tasks. Our research also differs from past work in that, to the best of our knowledge, there are no computational studies of perceptual learning for auditory tasks.

Here we model auditory perceptual learning across two tasks: temporal interval discrimination and frequency discrimination, as discussed in Wright and Sabin (2007). By modeling learning across several tasks our goal is to gain a better understanding of why learning does or does not occur under various training conditions. Our focus here is on modeling the daily limits of learning: it was observed in Wright and Sabin (2007) that training beyond some point in a single day does not yield extra learning. Our results suggest that limits on daily learning can be modeled by a decay of the memory of trials observed on the current day of practice. This decay is consistent with numerous studies of consolidation suggesting newly acquired information in a day begins in a volatile state, and is not made permanent until memories are consolidated (e.g. McGaugh, 2000).

## Human Data

This section reviews the human data and results originally described in Wright and Sabin (2007). In this paper, they examined how varying the number of training trials practiced per day affected learning over multiple days on two auditory discrimination tasks: frequency discrimination and temporal-interval discrimination. The basic question asked in the paper was “how much daily training is sufficient for learning to occur?” The set of relevant findings we model here is that extra trials practiced per day, past a certain point, do not appear to lead to any further learning.

During the experiments, subjects practiced either a temporal interval discrimination task or a frequency discrimination task for a single session each day of practice, for six days over no more than two weeks. Each task was a two inter-

val forced choice: on each trial participants must pick which of two stimuli is longer (higher) for the interval (frequency) discrimination task. The stimuli were adjusted adaptively as practice continues. As subjects do better, the difference between the standard (shorter) and comparison (longer) stimulus gets smaller. This is a common procedure used in psychophysics to find a performance threshold. The experiments consisted of a two-by-two design over number of trials in a day (360 or 900) and task type (frequency or interval). Each of the four conditions used a different set of participants. Further details of the training procedure can be found in Wright and Sabin (2007).

The data suggest there are important within-day limitations on human perceptual learning: extra practice past some point does not improve learning any further and insufficient practice in a day yields little to no learning across days. Further, the number of trials needed for learning is task dependent. Specifically, if a subject practiced the temporal interval task for 360 trials per day this yielded the same amount of learning as 900 trials per day. During the practice of frequency discrimination, 900 trials of practice produced significantly more learning than 360 trials. All the above observations were statistically verified. Details can be found in Wright and Sabin (2007).

Here our focus will be on modeling this first observed limit within a day: past a certain point no further trials within a day appear to yield further learning.

## Method

This section describes and justifies the basic principles of our model (which is evaluated in our Results section).

In terms of Marr's (1982) levels of analysis, we restrict ourselves largely to the informational level. When operating at this level we make no claims about what algorithm is used internally or how that algorithm is implemented in the human brain. Since the informational constraints are not yet fully understood for the modeled experiments, we believe this is an appropriate level of analysis for the time being.

Specifically, we utilize an ideal observer analysis (Geisler, 2003). The idea is to consider human performance in reference to an ideal observer, which processes information in a way that is 'optimal' in some sense. This can help to avoid conflation between two distinct types of limitations on human behavior. These are, respectively, informational and psychological limits. Informational limits are those limits that are inherent to the task: even if an observer were to be perfect they would still be subject to informational limits. An example of an informational limit would be noise in the input: any learner, no matter how smart, would have to deal with the problems introduced by noise. Psychological limits on the other hand are a product of resource limitations on the part of the observer: if the observer was 'smarter' they might be able to improve their behavior. An example of a psychological limit would be memory: with limited memory only so many units of information can be stored, but a smarter learner

would be able to store more, and so improve behavior.

Since any observer is subject to informational limits, we always assume these are present. Psychological limits are then only hypothesized as necessary: if a behavior could be explained solely in terms of informational limits, then no additional psychological limits would be hypothesized. Throughout our discussion we make a distinction between the ideal observer and the proposed psychological limits.

Based upon this principle we designed a system capable of modeling the observed limits on the amount of useful daily practice, as observed in Wright and Sabin (2007). We begin by describing the commitments we made regarding what information is available to humans when performing this task. We then describe an ideal observer model, and then identify the ways in which our model of human performance differs from the ideal observer.

## Input

The input to our model is consistent with the following properties, which are explained in more detail below. These choices represented a number of educated guesses as to the form of the information humans receive, based on psychophysical and physiological findings.

1. Differentiation along task relevant dimensions: e.g. 1 kHz is represented differently than 2 kHz.
2. Corruption by noise.
3. Range specificity: e.g. energy near 1 kHz is encoded separately from energy near 2 kHz.
4. Weber's law.

Each of these properties is based on many observations. Clearly the input is differentiated along task relevant dimensions: if there was no differentiation at all along a task relevant dimension, different stimuli of a task would appear the same to us. Second, there are many evident sources of noise to perceptual data, from noise in the world, noise during the transduction of sound to neural impulses, and noise in the nervous system itself. Range specificity is consistent with the narrow generalization patterns observed during perceptual learning tasks (e.g. Poggio et al., 1992; Fiorentini and Berardi, 1980; Wright and Zhang, 2009) and with the great multitude of physiological data suggesting that neurons are responsive to specific, limited ranges of stimuli (e.g. Brugge, 1992; De Valois and De Valois, 1980). Range specificity is distinct from differentiation: for instance a single source of information can differentiate between 1000 Hz and 200 Hz by using a single number, 1000 or 200, which would not be specific to a particular range; range specificity means that the sources of information (e.g. neurons) representing 1000 Hz and 200 Hz would be at least somewhat disjoint.

Weber's law—which states that the minimum discernible difference (or just noticeable difference) between stimuli along a particular dimension is proportional to the magnitude

of the stimuli along that dimension—has long been established as a useful rule of thumb for perceptual data (Moore, 2006).

In addition we make a number of simplifying assumptions. We assume that, prior to perceptual learning, the input has been correctly broken down into the various experimentally relevant units (i.e. each input to our model represents a single stimulus). How this happens in humans is not the focus of this modeling experiment. Our second assumption is that the dimensions of the stimulus are independent cues for the tasks in question, which is correct for the two tasks we consider.

Frequency and temporal interval are represented on a log scale. The frequency representation is found directly from the model described in (Wang and Shamma, 1994)<sup>1</sup>. Our interval representation is found based on a windowed autocorrelation of the stimulus onsets. Both of these choices yield a representation consistent with our above assumptions. The input to the observer is a vector  $\mathbf{x}$  of 228 terms: 128 features representing frequency and 100 features representing temporal interval. There are 128 bins for frequency because this is the resolution of the model from (Wang and Shamma, 1994). The number 100 for the interval representation was chosen arbitrarily. The observations made in the Results section did not change when this number was changed to 50 or 200.

We permute the input by an experimentally determined amount of noise specific to each dimension of the stimulus ( $\sigma_t^2$  for the interval noise and  $\sigma_f^2$  for the frequency noise). Note that since the representation is deterministic, when it is applied directly to an ideal observer it would always respond correctly. Choosing to represent all error in the system as input noise is conservative in the sense that the ideal observer will do more poorly under these conditions than if some of the error was modeled as output noise, for instance.

### Ideal Observer

We implement the ideal observer using a Bayesian approach to learning: a probabilistic model which is learned during the course of practice is used to determine the correct response on each practice trial. This model is not meant to be a psychologically plausible model of perceptual discrimination. It is an optimal decision maker for this task, whose performance can thus be used to identify in what ways humans are different from an optimal choice.

For a single trial, there are two stimuli, and each stimulus is encoded as a vector,  $\mathbf{x}$ , of 228 terms: 128 features for the frequency representation and 100 for the interval representation. Since we know that this input is permuted by Gaussian noise the likelihood of each stimulus type—the standard (or longer) and the comparison (or shorter)—can be modeled using a Normal distribution. We calculate the posterior model analytically by assuming a conjugate prior (Gelman, 2004). Learning and use of this model then follows a straightforward application of Bayes rule and conjugate priors, described be-

low.

Specifically the ideal observer learns a model of the standard (e.g. shorter) stimulus,  $S$ , and one for the comparison (e.g. longer) stimulus,  $C$  for each task. Each model is a multivariate Normal distribution, describing the probability of observing a given input vector  $\mathbf{x}$ . This distribution is specified by the mean vector  $\mu_S$  for the standard model and  $\mu_C$  for the comparison. Each mean has 228 terms (one for each frequency and interval value) and covariance matrix  $\Sigma_S$ , or  $\Sigma_C$  with 228 rows and columns. Hence, the probability of observing a given input vector, assuming it is the standard is as follows.

$$p(\mathbf{x}|\mu_S, \Sigma_S) \propto \exp [(\mathbf{x} - \mu_S)^T \Sigma_S^{-1} (\mathbf{x} - \mu_S)] \quad (1)$$

To learn the model of  $S$  and  $C$  the observer must be provided with examples of the standard and the comparison. These can be used to determine the probability of a given  $\mu_S$  and  $\Sigma_S$ , using Bayes rule. Below  $\mathbf{x}_t$  represents the example of the standard (shorter) stimulus observed at time  $t$ . On each practice trial, feedback is given to the observer after it responds, so on each trial the observer is provided with another example of both the standard and the comparison.

$$p(\mu_S, \Sigma_S | \mathbf{x}_1) \propto p(\mathbf{x}_1 | \mu_S, \Sigma_S) p(\mu_S, \Sigma_S) \quad (2)$$

$$p(\mu_S, \Sigma_S | \mathbf{x}_1, \mathbf{x}_2) \propto p(\mathbf{x}_2 | \mu_S, \Sigma_S) p(\mu_S, \Sigma_S | \mathbf{x}_1) \quad (3)$$

⋮

$$p(\mu_S, \Sigma_S | \mathbf{x}_t, \dots, \mathbf{x}_1) \propto p(\mathbf{x}_t | \mu_S, \Sigma_S) p(\mu_S, \Sigma_S | \mathbf{x}_{t-1}, \dots, \mathbf{x}_1) \quad (4)$$

Equation 2 requires that the prior probability  $p(\mu_S, \Sigma_S)$  be known, which we will discuss shortly. Subsequent equations show how an example  $\mathbf{x}_t$  updates the distribution of parameters for  $S$ . Given a set of training examples, the probability of  $\mathbf{x}$  for model  $S$  is defined as follows:

$$p(\mathbf{x}|S) = \iint p(\mathbf{x}|\mu_S, \Sigma_S) p(\mu_S, \Sigma_S | \mathbf{x}_t, \dots, \mathbf{x}_1) d\mu_S d\Sigma_S \quad (5)$$

Equation 5 can be calculated given that conjugate priors are used. Once  $p(\mathbf{x}|S)$  and  $p(\mathbf{x}|C)$  are known, Bayes rule can be used to find the probability that the model should indicate that the first (or second) stimulus is the longer of the two stimuli presented on a trial.

To use this Bayesian learner we must define the prior of the model ( $p(\mu, \Sigma)$ ), representing what people know before they practice the task. There are many deep questions that might be asked about what humans know about task before practice and how they know it. Here we choose a simple approach to selecting a prior: starting with a naive model (with mean vector 0, and an identity matrix for covariance) the learner is presented an experimentally determined number of trials of each task ( $N_t$  trials of the interval task, and  $N_f$  trials of the frequency task).

<sup>1</sup>An implementation of this model can be found at <http://www.isr.umd.edu/Labs/NSL/Register.htm>.

## Psychological Limits

We consider two modifications of the ideal observer described in the previous section to model psychological limits. The first is a direct result of the observation in (Wright and Sabin, 2007) that for these tasks people do not appear to learn within a day but only across days, hence our ‘daily’ model. The ‘daily’ model learns as per the ideal observer, but *responds* based only on data from previous days of practice, and not from the current day. This is used as a baseline model during our evaluation in the next section. Our second modification models the hypothesis that there is a daily limit on training: it does this by introducing a decay on the knowledge obtained from trials on the current day. The ‘decay’ model incorporates this limit, in addition to the limits of the ‘daily’ model. This proposed decay is a novel contribution of this paper in that it has not been considered as an explanation for the observed daily limit in these tasks before.

The decay in the model is implemented as follows. Given a new example,  $\mathbf{x}_{t+1}$  at trial  $t + 1$ , normally the model of the standard (or comparison) stimulus is updated according to Bayes rule in the following manner.

$$f_{t,d}(\mu, \Sigma \mid D_{t+1}, C) \propto p(\mathbf{x}_{t+1} \mid \mu, \Sigma) f_{t,d}(\mu, \Sigma \mid D_t) f_{T,d-1}(\mu, \Sigma \mid C) \quad (6)$$

In Equation 6, the function  $f_{t,d}$  is the distribution over stimulus parameters  $\mu$  and  $\Sigma$ , on trial  $t$  of day  $d$ .  $D_t$  represents all training examples observed for the current day, and  $C$  represents all examples observed on previous days (i.e. the consolidated information).  $T$  is the maximum number of trials observed in a day. This expresses the same relation expressed in Equation 4. However, with memory decay, this optimal update is changed to the following rule.

$$f_{t,d}(\mu, \Sigma \mid D_{t+1}, C) \propto p(\mathbf{x}_{t+1} \mid \mu, \Sigma) f_{t,d}(\mu, \Sigma \mid D_t)^{1-L} f_{T,d-1}(\mu, \Sigma \mid C) \quad (7)$$

Equation 7 means that memory decay occurs for trials observed on the current day. The distribution learned from a previous day of practice remains in the same state it was at the end of that day of practice (as determined by  $f_{T,d-1}$ ), including any decay that occurred on that day. This decay is a reasonable representation of loss of information within a day. If  $L = 0$  then the model is equivalent to the ‘daily’ model. If  $L = 1$  the daily practice has no effect on the model. Values between 1 and 0 represent a continuum between these two extreme conditions.

Note that it’s possible the decay should be over some shorter period of time, rather than including all trials within a day. For instance, it has been suggested that if a short nap is taken this has the same benefit as a night of sleep for purposes of perceptual learning (Mednick et al., 2003). This could easily be explained by our model by having  $D_t$  contain only those trials that occur after the last period of sleep, and  $C$  contain all other trials. However, this is beyond the scope of the experiments modeled in this paper.

## Results

Our hypothesis is that the observed daily limits on learning can be modeled as a decay of the memory of trials on the current day (while leaving memory of previous days’ trials untouched). We compared a computational model that had this hypothesized limit (the ‘decay’ model) to one that did not (the ‘daily’ model). To compare these models to human data we ran the same adaptive track blocks used in (Wright and Sabin, 2007) to determine thresholds. On each trial the original audio input was represented to the model and a response was given, and then feedback about the correct answer was used by the model to learn. This procedure was repeated 30 times, to simulate 30 different experimental subjects. This number was chosen to yield satisfactory statistical power for our analysis.

Results for the two models are discussed below. Figure 1 displays the results of these two models alongside human performance, as observed in (Wright and Sabin, 2007). From the graphs it appears that both models appear to fit the results well for the 360 trials/day interval discrimination condition and the 900 trials/day frequency discrimination condition. The decay model appears to also fit the data for the 900 trials/day interval discrimination condition better than the daily model.

Our statistical tests supported this observation. For each iteration, condition and day of a model we found the squared error to the mean human performance on that day. Table 1 shows the mean squared errors across conditions and models. Because the human and model data were qualitatively different in the 360 trial/day frequency condition we excluded it from the below analysis, since any differences between the two models in this condition will not be meaningful. A 3x2x6 ANOVA across conditions and models and within days of these squared errors showed a main effect across condition and model ( $p < 0.028$ ). A Tukey’s HSD test suggested that the decay model’s mean squared error was significantly less than the daily model’s mean square error ( $p < 0.014$ ).

	Interval		Frequency	
	360	900	360	900
<b>daily</b>	2.68(0.32)	3.40(0.37)	18.09(1.1)	1.24(0.11)
<b>decay</b>	2.77(0.29)	2.03(0.19)	24.60(1.2)	1.19(0.14)

Table 1: Mean squared errors for the daily and decay model. Errors are the difference between a model threshold and the mean for the human data on a given day and condition. Numbers in parenthesis indicated standard errors.

Model parameters (which determined noise and prior knowledge) were adjusted so that the daily model matched human performance on day 1 and day 6 of all conditions except the 360 trials/day frequency condition, using the optimization algorithm described in Huyer and Neumaier (2008). These conditions were chosen because this was where learning appeared to occur. Since the noise of the model strongly influences the final performance of our model on day 6 (after learning), it should be fit to those conditions where learning

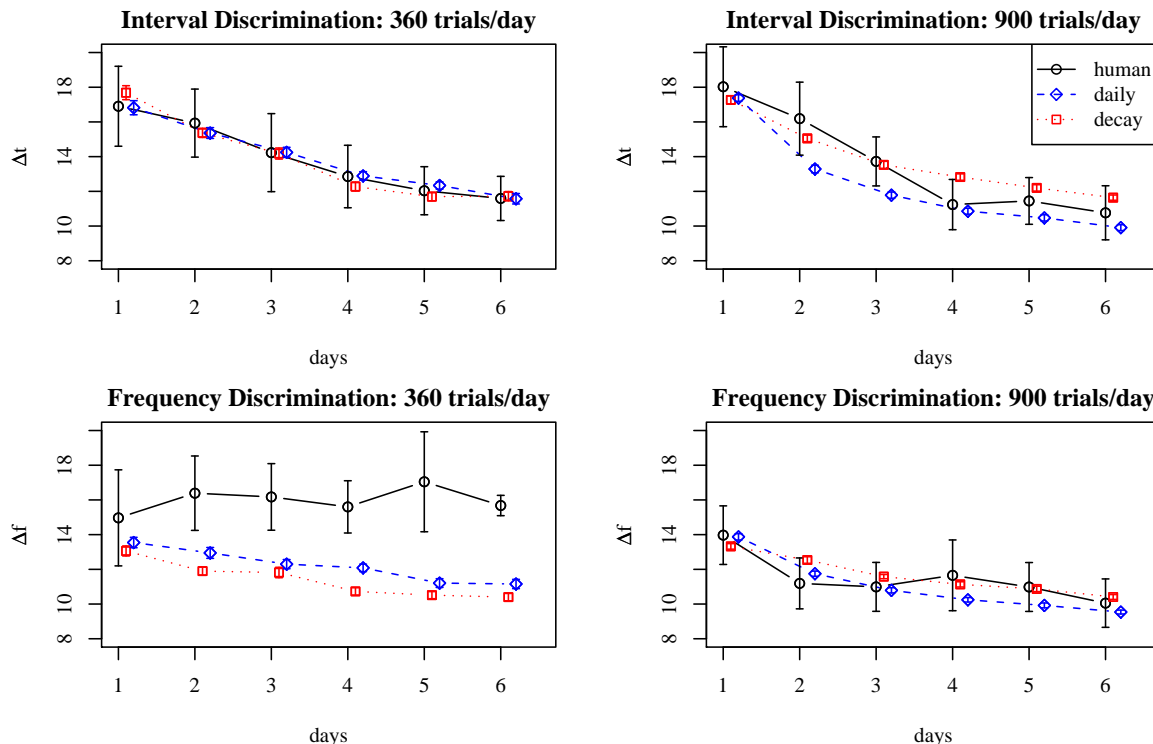


Figure 1: Results for ‘daily’ and ‘decay’ models compared to human performance. Results are averaged across 30 runs of each model.  $\Delta f$  represents the difference between the standard (lower) and comparison (higher) frequency stimuli for the frequency task, and  $\Delta t$  the difference between the standard (shorter) and comparison (longer) stimuli for the interval task. The adaptive track method used finds the 79% accuracy of a subject or model. Lower delta’s indicate that the human participants are performing better. A model is accurately predicting the human data if its curve is closer to the human curves. Bars indicate standard errors.

appears to occur. The parameters for prior knowledge are dependent on this noise and so we fit it jointly and under the same conditions as the noise. For reasons that will become clear below we also matched this data to human performance on day 2 of the 900 trials/day interval discrimination task.

An analogous procedure was used for the decay model except that the decay parameter ( $L$ ) was also adjusted, and fit to the same days as above. The data was fit to day 2 for the 900 trials/day interval. This single day was chosen so as to be minimal (to avoid overfitting) and such that it was a place where  $L$  might cause an observable change in the results. This same day was used for the daily model above so that both procedures had access to the same information. All parameters were selected so as to maximize the posterior probability of the selected days given the human thresholds (assuming thresholds on a day are Normally distributed, which is consistent with the analysis in Wright and Sabin (2007)).

## Discussion & Conclusions

In this paper we evaluated a model of learning across two simple auditory tasks. Our goals differed from that of previous work (e.g. Poggio et al., 1992; Petrov et al., 2005; Jacobs, 2009) in that we considered auditory tasks rather than

visual tasks, and in that we considered a single model that could explain results across several tasks. To the best of our knowledge, ours is the first computational model of auditory perpetual learning.

Our contributions in this paper were to show that our ‘daily’ model could accurately model two of the four considered experimental conditions and that our ‘decay’ model (which included a decay of memory for the trials observed on current days) could model an additional condition (900 trials/day of interval discrimination). This result suggests that the minimal difference in learning for this condition and the 360 trials/day of interval discrimination could be caused by memory loss.

Modeling this condition using memory decay is consistent with numerous studies of consolidation suggesting newly acquired information begins in a volatile state, and is not made permanent until consolidation occurs after practice is complete (McGaugh, 2000). In cases where consolidation is interfered with, perhaps what happens is that the memory of observed trials on a task decays before it can be stored in long term memory. The 900 trial/day interval discrimination condition would then represent an intermediate case where consolidation has yet to occur (perhaps because practice is still

ongoing), and hence memory decay degrades part of what has been learned. Once practice is complete consolidation can commence given that no other interfering effects occur.

The model presented here does not explain one of the experimental conditions we considered (the condition with 360 a trials of frequency discrimination a day). In this condition people did not appear to learn but our model did, suggesting that the human results cannot be explained simply by the fact that fewer trials were observed, which is consistent with the observations made in Wright and Sabin (2007). We have considered several possible factors that might explain this condition, but as of yet, no factor we have considered can explain both the 360 trial interval discrimination task and the 360 trial frequency discrimination task using a single parameter. Any model using a different parameter per condition would be meaningless in that any such model would fit the data. This suggests to us that more perceptual learning tasks must be considered before a meaningful model for this condition and others like it can be proposed, and is a goal of future work. In the future, it is also our plan to consider conditions where people practice several tasks at once, to help us understand why learning does or does not occur, such as in (Banai et al., 2009).

This paper thus represents a first step toward developing a model that can explain learning across a number of perceptual learning tasks, rather than modeling behavior on a single task. Such a model must consider more constraints than one that doesn't, which can help provide a better understanding of how and when perceptual learning occurs and why.

### Acknowledgements

We'd especially like to give thanks to Beverly Wright for her helpful comments, and to Andy Sabin for his help providing access to the data from Wright and Sabin (2007). This research was supported, in part by Northwestern University's Cognitive Science program and by US National Science Foundation grant 0643752.

### References

- K. Banai, JA Ortiz, JD Oppenheimer, and BA Wright. Learning two things at once: differential constraints on the acquisition and consolidation of perceptual learning. *Neuroscience*, 2009.
- J Brugge. *The Mammalian Auditory Pathway: Neuroanatomy*, volume I, chapter An overview of central auditory processing., pages 1–22. Springer Verlag, Wiesbaden, Germany, 1992.
- RL De Valois and KK De Valois. Spatial vision. *Annual Review of Psychology*, 31(1):309–341, 1980.
- Laurent Demany. Perceptual learning in frequency discrimination. *The Journal of the Acoustical Society of America*, 78(3):1118–1120, 1985. doi: 10.1121/1.393034. URL <http://link.aip.org/link/?JAS/78/1118/1>.
- M. Fahle and T. Poggio. *Perceptual learning*. MIT Press Cambridge, MA., Cambridge, MA, USA, 2002.
- A. Fiorentini and N. Berardi. Perceptual learning specific for orientation and spatial frequency. *Nature*, 287:43–44, 1980.
- A. Gelman. *Bayesian data analysis*. CRC press, 2004.
- E.A. Hayes, C.M. Warrier, T.G. Nicol, S.G. Zecker, and N. Kraus. Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, 114(4):673–684, 2003.
- W. Huyer and A. Neumaier. SNOBFIT—Stable Noisy Optimization by Branch and Fit. *ACM Transactions on Mathematical Software (TOMS)*, 35(2):9, 2008.
- RA Jacobs. Adaptive precision pooling of model neuron activities predicts the efficiency of human visual learning. *Journal of Vision*, 9(4):22, 2009.
- P.J. Kellman and P. Garrigan. Perceptual learning and human expertise. *Physics of Life Reviews*, 2008.
- P.J. Kellman, C. Massey, Z. Roth, T. Burke, J. Zucker, A. Sawa, K.E. Aguero, and J.A. Wise. Perceptual learning and the technology of expertise Studies in fraction learning and algebra. *Pragmatics & Cognition*, 16(2):356–405, 2008.
- D. Marr. *Vision: A Computational Approach*. Freeman & Co., San Francisco, 1982.
- James L. McGaugh. Memory—a century of consolidation. *Science*, 287(5451):248–251, 1 2000.
- S. Mednick, K. Nakayama, and R. Stickgold. Sleep-dependent learning: a nap is as good as a night. *Nature Neuroscience*, 6(7):697–698, 2003.
- Brian C. J. Moore. *An Introduction to the Psychology of Hearing*. Elsevier, London, UK, 5th edition, 2006.
- D. Norris, J.M. McQueen, and A. Cutler. Perceptual learning in speech. *Cognitive Psychology*, 47(2):204–238, 2003.
- A.A. Petrov, B.A. Doshier, and Z. Lu. The Dynamics of Perceptual Learning: An Incremental Reweighting Model. *Psychological Review*, 112(4):715, 2005.
- Tomaso Poggio, Manfred Fahle, and Shimon Edelman. Fast perceptual learning in visual hyperacuity. *Science*, 256(5059):1018–1021, 1992. URL <http://www.jstor.org/stable/2877128>.
- Kusan Wang and Shihab Shamma. Self-normalization and noise-robustness in early auditory representations. *IEEE Trans. On Speech and Audio Processing*, 2(3):421–435, 1994.
- B.A. Wright and Y. Zhang. A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515):301, 2009.
- B.A. Wright, D.V. Buonomano, H.W. Mahncke, and M.M. Merzenich. Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, 17(10):3956–3963, 1997.
- Beverly Wright and Andrew Sabin. Perceptual learning: how much daily training is enough? *Experimental Brain Research*, 180(4):727–736, 07 2007.