UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation

Permalink

https://escholarship.org/uc/item/9v45w71z

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN

1069-7977

Authors

Kleinschmidt, Dave F.
Jaeger, T. Florian

Publication Date

2012

Peer reviewed

A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation

Dave Kleinschmidt[1] and **T. Florian Jaeger**[1,2]
{dkleinschmidt, fjeager} @ bcs.rochester.edu
[1]Department of Brain and Cognitive Sciences and [2]Department of Computer Science,
University of Rochester, Rochester, NY, 14607 USA

Abstract

We have previously proposed that incremental belief updating can provide a unified account of the effect of cumulative exposure on phonetic recalibration and selective adaptation (Kleinschmidt & Jaeger, 2011). This model predicts that these are not two distinct phenomena but rather two points on a continuum. We investigate that prediction here using adaptor stimuli intermediate between those which induce recalibration and selective adaptation, and find that the quantitative predictions of the model fit the data well. We also demonstrate that with the proper controls, Mechanical Turk provides a suitable online platform for speech perception experiments.

Keywords: Phonetic adaptation; Speech perception; Recalibration; Selective adaptation; Bayesian models; Mechanical Turk

Introduction

Language is fraught with variability, and such variability is a particular problem in the mapping from acoustic cues to phonetic categories. Some of this variability is simply random noise, from imprecision in the production and perception systems, but much of it is systematic, reflecting real differences in the way that individual speakers realize phonetic categories.

One of the ways in which the human speech perception system deals with such systematic variability is through rapid recalibration of phonetic categories in response to novel accents (Bradlow & Bent, 2008; Kraljic, Brennan, & Samuel, 2008) and unusual pronunciations of particular phonetic categories (Norris, McQueen, & Cutler, 2003; Vroomen, Linden, Gelder, & Bertelson, 2007). Through repeated exposure to a sound which is acoustically ambiguous between, say, /b/ and /d/, but which is disambiguated via visual or lexical information as a /b/, the listener's categorization behavior will change: they will classify more items on a /b/-to-/d/ continuum as /b/.

Such recalibration is often contrasted with selective phonetic adaptation, where repeated exposure to an *prototypical* /b/ results in exactly the opposite effect on the listener's categorization behavior: they will classify fewer items on the continuum as /b/ (Eimas & Corbit, 1973; Samuel, 1986; Vroomen, Linden, Keetels, Gelder, & Bertelson, 2004; Vroomen et al., 2007). Selective adaptation and recalibration are generally held to be qualitatively distinct, for a number of reasons.

First, recalibration results from repeated exposure to an acoustically ambiguous stimulus, while selective adaptation relies on repeated exposure to an unambiguous stimulus, and

is in fact stronger for more prototypical stimuli (Miller, Connine, Schermer, & Kluender, 1983). Second, despite the fact that both come from repeated exposure to a phonetic category member, they have opposite effects on that category.

Third, selective adaptation and recalibration are thought to reflect different levels of processing. On the one hand, selective adaptation is thought to generally reflect low-level, perceptual processing (Samuel, 1986). On the other, phonetic recalibration (or "perceptual learning for speech", Norris et al., 2003) seems to reflect reorganization of the phonetic categories themselves (Clarke-Davidson, Luce, & Sawusch, 2008; Maye, Aslin, & Tanenhaus, 2008), is long-lasting (Eisner & McQueen, 2006), and in some cases generalizes to new speakers and phonetic contrasts (Kraljic & Samuel, 2006).

Fourth, the amount of cumulative exposure has a dramatically different effect on selective adaptation and recalibration (Vroomen et al., 2007). Selective adaptation grows stronger with more exposure, while recalibration becomes weaker after an initial rise, to the extent that it is eventually extinguished (Figure 2, left).

Despite these differences, selective adaptation and recalibration have something in common: they are reactions to unusual distributions of acoustic-phonetic cues. Listeners are exquisitely sensitive to such distributional information (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Maye, Werker, & Gerken, 2002), and we propose that incremental belief updating might provide a unified account of a range of phonetic adaptation phenomena, possibly including selective adaptation (Kleinschmidt & Jaeger, 2011).

This approach suggests that at least in some cases, selective adaptation and perceptual recalibration are better understood as part of a continuum of phonetic adaptation effects, rather than as qualitatively distinct processes, which raises the question of what sort of adaptation effect intermediate adaptors will produce. In this paper, we investigate the quantitative predictions about such intermediate adaptors made by our belief-updating model.

First, we describe how incremental belief updating provides a unifying account of the effects of cumulative exposure on recalibration and selective adaptation (Kleinschmidt & Jaeger, 2011). This model is used to derive quantitative predictions for the effect of cumulative exposure to audio-visual speech stimuli on a continuum between acoustically ambiguous and prototypical. These predictions are tested against human behavior. Finally, we conclude with a discus-
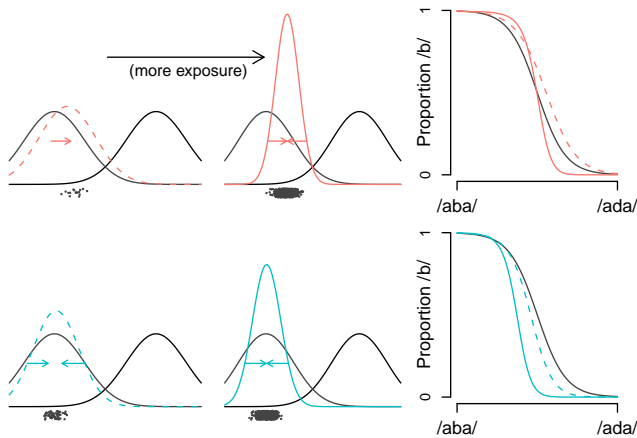
Figure 1: Qualitative predictions of the Bayesian belief updating model for phonetic recalibration (top) and selective adaptation (bottom). Left panels show changes in mean/variance after a small number of adaptor trials, and middle panels show the effect of further exposure. Note that while the adaptor for recalibration is acoustically ambiguous, it is disambiguated (by means of, e.g., a visual /aba/), and thus the percept is biased towards the /b/ prototype. Right panels show corresponding identification curves.

sion of future directions for both the computational modeling and behavioral work.

Model and predictions

Speech perception can be considered a process of statistical inference: given a set of acoustic-phonetic cues, what is the most likely intended phonetic category (Sonderegger & Yu, 2010; Feldman, Griffiths, & Morgan, 2009; Kleinschmidt & Jaeger, 2011)? This inference critically depends on accurate beliefs about the distributions of acoustic-phonetic cues for each category. If the listener's expectations about what a /b/ sounds like do not match what the speaker produces then the listener is at rist of mistaking the speaker's intended phonetic category. Thus, a rational listener must update their expectations based on experience.

Figure 1 shows how such incremental belief updating explains the qualitative differences between recalibration (top) and selective adaptation (bottom) based on the respective distributions of acoustic cues (Kleinschmidt & Jaeger, 2011). Phonetic categories are naturally "wide", with random variability inherent in perception and production processes. Selective adaptation effects are captured by the fact that in the prototypical-/b/ condition, the unusually tight clustering of acoustic cues from the repeated adaptor stimulus causes the /b/ category to shrink, resulting in fewer /b/ responses to test stimuli from the middle of the continuum.

For recalibration, the initial rise in /b/ responses after exposure to the acoustically ambiguous (but visually or lexically disambiguated) /b/ adaptor is due to a shift in the *mean* of the /b/ category, towards the middle of the continuum. With more exposure, however, the tight clustering again causes the /b/ category to contract. Because the acoustically ambiguous adaptor is biased towards the prototypical /b/ by the disambiguating (visual or lexical) cue (by a McGurk effect or perceptual magnet effect, etc.), the shrinking of the /b/ category means that, like with selective adaptation, it pulls away from the middle of the continuum, resulting in an eventual decrease in /b/ responses.

This belief-updating model treats the behavioral phenomena of selective adaptation and recalibration as two points on a continuum of phonetic adaptation effects, which is determined by the statistics of the adaptor stimuli. One way of describing these statistics is by the mean and the variance of the adaptor stimuli, and our model makes quantitative predictions about the whole range of adaptors, from prototypical adaptors (which produce selective adaptation) to auditorily ambiguous adaptors (which produce phonetic recalibration), as well as intermediate adaptors, where the adaptor is neither fully auditorily ambiguous nor prototypical.

Specifically, phonetic adaptation with such intermediate audio-visual adaptors should be intermediate in time course and in direction. One simple way of characterizing the time course of phonetic adaptation is the number of exposures at which the direction of the effect of /b/ exposure switches from more /b/ classification to less /b/. In Vroomen et al. (2007, replotted in Figure 2, left), this "switch point" for the ambiguous condition is essentially the end of one full block: /b/-classification is greater after ambiguous /b/ (solid lines) than after ambiguous /d/ exposure (dashed lines) until the very end of the block, at 256 trials, where the two become indistinguishable again. For the prototypical condition, there is less /b/ classification after /b/ exposure from nearly the first adaptor repetition, and so can we say that the switch point is at the beginning of the block. For intermediate adaptors, this switch point will be between the point for fully ambiguous and fully prototypical adaptors, closer to the ambiguous switch point for more ambiguous adaptors and closer to the unambiguous point for less ambiguous adaptors (Figure 3, black curves with arrows showing switch points).

Model specification

We refer the reader to Kleinschmidt and Jaeger (2011) for full details of the model. This model treats phonetic categories as Normal distributions over acoustic-phonetic cues, $p(x_i|c_i) \sim \mathcal{N}(\mu_{c_i}, \sigma_{c_i}^2)$. The listener's beliefs about these categories are captured by Normal-Gamma probability distributions over the category means and variances, $p(\mu_c, \sigma_c^2)$. After exposure to adapting stimuli X from category c, the listener's beliefs are updated via Bayes rule, $p(\mu_c, \sigma_c^2|X, c) \propto p(X|\mu_c, \sigma_c^2)p(\mu_c, \sigma_c^2)$. This finds the best compromise between the listener's prior beliefs and the mean and variance that is most consistent with the data.

The probability of a /b/ response to a stimulus x, $p(b|x)$, depends on the likelihood of x under each category $p(x|b)$ and
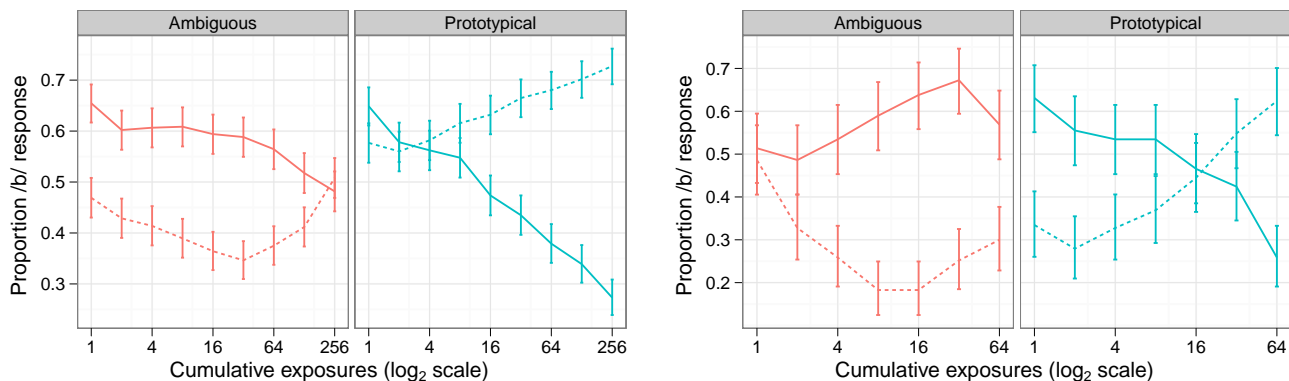
Figure 2: Results from Vroomen et al. (2007), left, and Experiment 1, right: More /b/-classification after ambiguous /b/ exposure (solid lines, left panels) than /d/ exposure (dashed lines). After prototypical-/b/ exposure (solid lines, right panels), less /b/-labeling.

$p(x|\text{d})$, as well as the prior probability of /b/ and /d/:

$$p(\text{b}|x) = \frac{p(x|\text{b})p(\text{b})}{p(x|\text{b})p(\text{b}) + p(x|\text{d})p(\text{d})}$$

The predictions of the model depend not only on the mean and variance of the adaptor stimuli (the likelihood function) but also the prior beliefs about the category means and variances. While each subject's most likely means and variances can be determined based on their pre-test classification data, the confidence in these beliefs must be treated as free parameters (Kleinschmidt & Jaeger, 2011).

Here we incorporate a few minor improvements over our original model. Instead of fitting the after-effect measure used by Vroomen et al. (2007) (the difference between /b/ and /d/ exposure), here the model is fit to the /b/-classification rate curves for each individual condition. This requires accounting for asymmetry in the prior probability of /aba/ vs. /ada/, which was fixed to $p(\text{b}) = 0.7$, based on the English biphone probabilities of /aba/ and /ada/ (Vitevitch & Luce, 2004).[1] The free parameters are fit via MCMC sampling (rather than a grid-search) based on the binomial likelihood of the responses (using a weakly-informative prior).

We next describe experiments which test the predictions of this model. These are extensions of the paradigm of Vroomen et al. (2007), which we have adapted to run on a novel web-based platform. The first experiment replicates Vroomen et al. (2007) to validate this platform, and the second investigates the effect of cumulative exposure on adaptation to intermediate adaptors.

Experiment 1

Our first experiment is a replication of Vroomen et al. (2007). This is important both to verify the findings of Vroomen et

al. (2007) with English speakers but also to verify our web-based platform for running audio-visual phonetic adaptation studies.

Participants

24 participants were recruited and run via Amazon's Mechanical Turk service. An additional 4 subjects were excluded by the criteria described below.

Procedure

Our procedure is adapted from Vroomen et al. (2007), and our stimuli are identical in order to maximize similarity to the original study for purposes of replication (we thank Jean Vroomen for graciously providing them). These stimuli consist of a 9-item /aba/-to-/ada/ continuum synthesized by manipulating F2 locus in equal steps on a log-frequency scale. Audio-visual adaptor stimuli were created by combining a video of a speaker producing /aba/ or /ada/ with one continuum item, according to the condition.

Subjects first did a pre-test and calibration phase, where they classified items from the audio continuum as /b/ or /d/. The point of maximum ambiguity was found via logistic regression for each subject. Subjects who mis-classified any of the two most prototypical stimuli on each end (less than 70% accuracy) were excluded, as were any subjects whose maximally ambiguous stimulus was not one of the three middle continuum items 4, 5, or 6. Three subjects (11%) were excluded by these criteria and replaced.

During the main phase, in each of four blocks subjects saw 64 repetitions of an adaptor stimulus. This stimulus was visually either /b/ or /d/, and auditorily either a *prototypical*, unambiguous rendition of the visual category (continuum item 1 or 9, respectively), or auditorily *ambiguous* (that subject's most ambiguous continuum item as determined during pre-test). These four blocks were presented in latin-square order across subjects. Audio-only test stimuli were interspersed throughout each block in sets of six (two repetitions of each

---

[1] When the prior probability of /b/ vs. /d/ is treated as a free parameter, the best-fitting value is 63% chance of /b/, very close to the corpus-derived value.

of the three most ambiguous stimuli) after 1, 2, 4, 8, 16, 32, and 64 exposures.

To ensure that subjects were paying attention to the videos, catch trials were randomly interspersed, where a small white dot flashed for one frame below the nose of the speaker. Subjects who failed to respond to at least 80% of these trials (or missed more than 50% in one block) were excluded; only one subject was excluded by these criteria and replaced. Besides this subject, accuracy was 95% (compared to 93% in Vroomen et al., 2007).

Results

Because subjects were not run in the laboratory, there is some uncertainty about the conditions under which our subjects are doing the task. As one control for this, we compared their performance on the pre-test calibration task with the performance of the subjects in Vroomen et al. (2007) using a logistic mixed-effects regression model, with predictors for stimulus and group (web vs. lab) and with the maximum random effects structure justified by the data. Web and lab subjects did not differ in the slope of their category boundaries ($\beta = 0.07, \text{SE} = 0.14, p = 0.60$). The category boundary (point of maximum ambiguity) was on average about one half to one continuum step closer to the /b/ end of the continuum compared to the lab subjects ($\beta = -1.15, \text{SE} = 0.27, p < 10^{-4}$).

Our results from the main phase replicate Vroomen et al. (2007) as well (Figure 2). In the prototypical (selective adaptation) condition, subjects classify fewer test items as /b/ after /b/ exposure, and this effect grows stronger with further exposure. In the ambiguous (recalibration) condition they classify *more* as /b/ after /b/ exposure, but this effect eventually begins to weaken. Because our subjects only saw 64 exposure trials in each block, versus 256 in Vroomen et al. (2007), recalibration does not completely dissipate by the end of the block, but it does begin to weaken.

The most notable difference between our findings and those of Vroomen et al. (2007) is that in the prototypical, selective adaptation condition of our experiment, /b/-exposure initially produces *more* /b/-classification (as in the ambiguous, recalibration condition). There are two possible explanations for this. First, the "prototypical" stimuli we use correspond to natural productions of /aba/ and /ada/ by *Dutch* speakers, and may not be fully prototypical to English listeners, which, as we will see in Experiment 2, can produce recalibration-like effects with small amounts of exposure. Second, adaptation could carry over from one block to the next, producing baseline shifts. Such carry-over effects are observed in the data of Vroomen et al. (2007) as well, but are effectively confined to earlier blocks of their experiment (each of their subjects did a total of 16 blocks, versus only four here) and thus do not show up as strongly in the averaged data. In addition, at the end of the ambiguous (recalibration) blocks there is a bigger difference between /b/ and /d/ exposure in our data than in Vroomen's et al., which would magnify any carry-over effects on the prototypical blocks (which are preceded two-thirds of the time by an ambiguous block).

Experiment 2

The second experiment tests the predictions of our belief-updating model about the effect of cumulative exposure to adaptor stimuli intermediate between the prototypical and auditorily ambiguous stimuli used in the previous experiment. This also serves as another replication of Vroomen et al. (2007) using our web-based approach.

Participants

A total of 126 participants were recruited, run, and paid through Amazon's Mechanical Turk service as in Experiment 1. Six (5%) were excluded for poor catch trial performance, eight (6%) for misclassifying too many prototypical pre-test stimuli, and 18 (14%) for unusual category boundaries (as above).[2] After these exclusions, 94 subjects remained.

Procedure

The procedure was similar to that from Experiment 1, except that there were four ambiguity conditions rather than two, and they were distributed between subjects rather than within. In addition to the prototypical (selective adaptation) and ambiguous (recalibration) conditions from Experiment 1, there were two intermediate conditions. In the intermediate-ambiguous condition, the audio component was one step towards the visual category prototype from the maximally ambiguous stimulus, and in the intermediate-prototypical condition it was two steps.

Each subject performed two blocks: one /aba/ and one /ada/ of one of the four ambiguity conditions, in order to minimize the effects of carry-over between blocks. Each block was 128 exposure trials, with test trials as in Experiment 1.

Results

Figure 3 shows the results from all four conditions. As predicted by the belief-updating model, as the auditory component of the adaptor is less and less ambiguous, the switch point between recalibration-like and selective adaptation-like behavior moves closer and closer to the beginning of the block (Figure 4).

The quantitative predictions of the model are also a good match for the observed data. The predictions of the model are shown in black. These predictions correspond to the free parameters which best fit the ambiguous and unambiguous conditions. As in Kleinschmidt and Jaeger (2011), the model predictions provide a good fit for the behavior in the ambiguous and unambiguous blocks ($r^2 = 0.67$). More importantly, without refitting, the model predictions match just as well on the two intermediate conditions ($r^2 = 0.66$). That is, the model does not merely fit but quantitatively *predicts* the time course of adaptation to the intermediate adaptors.

---

[2]The large number of subjects excluded for unusual category boundaries could be attributed either to variability in listening equipment and environment, or to the fact that our listeners are English speakers but the continuum was generated based on Dutch productions of /aba/ and /ada/.
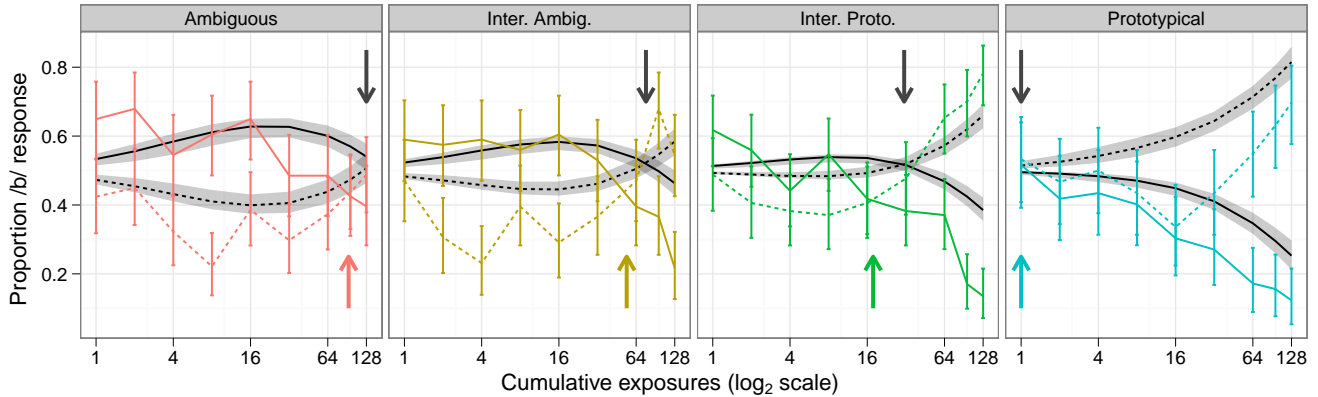
Figure 3: Model predictions and results from each subject's first block. Left to right shows fully ambiguous, intermediate-ambiguous, intermediate-prototypical, and prototypical. Solid lines are for visual-b exposures and dashed are visual-d. Colored lines are average /b/ response rate, and error bars are 95% confidence intervals on the mean. Black lines are model predictions based on best-fitting parameters to only ambiguous and unambiguous blocks, and shaded regions are 95% confidence intervals based on MCMC sampling of model free parameters with a weakly informative prior.
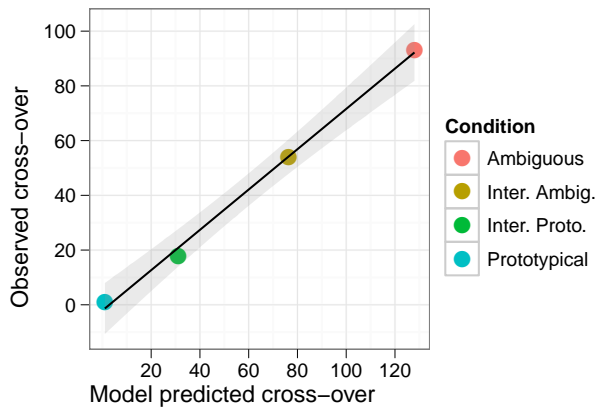

Figure 4: Predicted and actual recalibration-to-selective-adaptation switch points. The x-coordinate of each point is the predicted switch point (black curves/arrows in Figure 3), and the y-coordinate is the observed point (colored curves/arrows).

Discussion and conclusion

We have proposed that incremental belief-updating can provide a unified account of selective adaptation and phonetic recalibration, and have tested a model based on that principle against existing data on the effect of cumulative exposure on these two effects from Dutch listeners (Kleinschmidt & Jaeger, 2011). In this study we have first replicated this data (Vroomen et al., 2007) with American English listeners, using a novel, web-based platform, and second, extended the original study to test a key prediction of the model that existing data does not address.

It is generally argued that selective adaptation and recal-

ibration are qualitatively distinct effects (e.g., Vroomen et al., 2004). However, our model suggests that they both result from adaptation to changes in the statistics of the linguistic environment, and can be understood as two points on a continuum of phonetic adaptation phenomena. The data we present here shows that intermediate audio-visual adaptors produce adaptation which is intermediate in time course and direction in a way that is quantitatively consistent with the model predictions.

There is, of course, a large body of work which highlights the differences between selective adaptation and recalibration, and this study addresses only one such difference. We cannot, and do not, on the basis of this study claim that selective adaptation and recalibration *are*, in a mechanistic sense, the result of the same process. Rather, our approach offers some insight into these seemingly distinct adaptation phenomena by considering how the two behavioral patterns are related to the statistics of the linguistic environment.

Consider the alternative explanation of our results in terms of separate processes of selective adaptation (a decrease in /b/ responses after /b/ exposure) and recalibration (an increase in /b/ responses after /b/ exposure). Selective adaptation is stronger for more prototypical adaptors (Samuel, 1982; Miller et al., 1983), and so more the more selective adaptation-like time course of adaptation with increasingly prototypical adaptors might be explained simply by an interpolation between the time courses of recalibration and selective adaptation. However, this explanation leaves unanswered the question of why selective adaptation should be involved at all in recalibration, especially considering the fact that selective adaptation has been claimed to only be a function of the *audio* component of an AV stimulus (Saldaña & Rosenblum, 1994) and there is little-to-no selective adaptation observed for boundary stimuli (Miller et al., 1983).

Treating selective adaptation and recalibration as rational reactions to changes in the linguistic environment makes the intuitive idea about interpolating between the two time courses precise and quantitative, but more importantly *why*. As the listener encounters different speakers and different listening conditions, their linguistic environment is constantly changing, often in unpredictable ways. The listener needs to be able to cope with such changes and so must update their beliefs (and behavior) in response to their recent experience. However, while the reasons for such adaptation may be the same, the mechanisms by which it occurs could differ, with recalibration resulting, mechanistically, from higher-level language processing and selective adaptation occurring at a lower, perceptual level (Samuel, 1986).

The major goal of this work is to formulate a learning model for phonetic adaptation and use that model to guide experimental investigation of the whole range of phonetic adaptation phenomena. While this program emphasizes the similarities between different sorts of adaptation, it also aims to uncover even finer-grained empirical distinctions, in part to further the use of adaptation as a window on language processing, generally.

The model we have evaluated here is a first step and makes many simplifying assumptions. Still, it is a Bayesian model which, via data-determined priors and known processes—like sensitivity to category variance and the McGurk effect—provides a unified theoretical explanation for a puzzling interaction between the behavioral effects of selective adaptation and phonetic recalibration. More importantly, it makes fine-grained and testable predictions about the time course of behavior, some of which we have verified here.

Future work will focus on exploring the relationship between this rational model and boundedly rational implementations like exemplar models and particle filters. We will also, via structured, hierarchical extensions of this simple model, focus on resolving the apparent conflict between rapid adaptation to new speakers and the long-term stability of the linguistic system. Finally, and most importantly, we will use the predictions of these models to continue to systematically probe the space of phonetic adaptation behavior in order to inform and constrain the modeling work.

References

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–29.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, *70*(4), 604–618.

Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. a. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–9.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99–109.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–3.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, *116*(4), 752–82.

Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *2nd acl workshop on cognitive modeling and computational linguistics*.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262–8.

Maye, J., Aslin, R., & Tanenhaus, M. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science: A Multidisciplinary Journal*, *32*(3), 543–562.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–11.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *The Journal of the Acoustical Society of America*, *73*(6), 2124–33.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America*, *95*(6), 3658–61.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & psychophysics*, *31*(4), 307–14.

Samuel, A. G. (1986). Red herring detectors and speech perception: in defense of selective adaptation. *Cognitive psychology*, *18*(4), 452–99.

Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 375–380).

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481–7.

Vroomen, J., Linden, S. van, Gelder, B. de, & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–7.

Vroomen, J., Linden, S. van, Keetels, M., Gelder, B. de, & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: dissipation. *Speech Communication*, *44*(1-4), 55–61.