UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Modeling Adaptation to a Novel Accent

Permalink
https://escholarship.org/uc/item/2vv3m3dm

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Hitczenko, Kasia
Feldman, Naomi H.

Publication Date
2016

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Modeling Adaptation to a Novel Accent

Kasia Hitczenko[1] (khit@umd.edu)
Naomi H. Feldman[1,2] (nhf@umd.edu)
[1]Department of Linguistics and [2]Institute for Advanced Computer Studies
1401 Marie Mount Hall, University of Maryland, College Park, MD 20742

Abstract

Listeners quickly adapt to novel accents. There are three main hypotheses for how they do so. Some suggest that listeners expand their phonetic categories, allowing more variability in how a sound is pronounced. Others argue that listeners shift their categories instead, only accepting deviations consistent with the accent. A third hypothesis is that listeners both shift and expand their categories. Most work has supported the category expansion hypotheses, with the key exception of Maye et al. (2008) who argued for a shifting strategy. Here, we apply the ideal adaptor model from Kleinschmidt & Jaeger (2015) to reexamine what conclusions can be drawn from their data. We compare adaptation models in which categories are shifted, expanded, or both shifted and expanded. We show that models involving expansion can explain the data as well as, if not better than, the shift model, in contrast to what has been previously concluded from these data.

Keywords: accent adaptation; speech perception

Speech is highly variable. A given sound or word can be pronounced differently depending on the context, the speaker's accent or gender, and the discourse situation, among many other factors. In order for successful communication to take place, people's speech recognition system must adapt to this variability. Previous work has shown that listeners adapt to unfamiliar speech quickly. Although unfamiliar speech results in an initial processing cost, this cost generally disappears after a few minutes of exposure (Clarke & Garrett, 2004).

How do listeners adjust their sound categories in response to newly introduced variability? To address this question, we consider one example of adaptation to variable speech: adult perception of accented speech. In accented speech perception, the variability stems from the fact that the speaker and listener have different accents, so the speaker's pronunciation can differ, often substantially, from what the listener expects.

There are three main hypotheses about how individuals adapt their sound categories in response to accented speech. The *Expand* hypothesis suggests that individuals relax their categories, allowing more flexibility in how a particular category is produced. The *Shift* hypothesis suggests that individuals shift their categories, only allowing more flexibility in the direction of the accent they heard. Finally, the *Shift and Expand* hypothesis suggests that individuals do both: they allow more flexibility in how a given sound is produced, but do so around a shifted mean. Most work has found evidence that category expansion is involved in adaptation, either with a shift (*Shift and Expand*) or without (*Expand*).

Evidence for the *Expand* hypothesis comes primarily from experiments in which adults were exposed to an unfamiliar accent and, at test, had adapted to changes that were not present during the initial exposure (Schmale et al., 2012). In other words, they accepted variability that they had no evidence for. For example, listeners who were exposed to an accent in which back vowels were lowered accepted items that were raised versions of English words as words in a lexical decision task, despite not being familiarized with instances of this kind (Weatherholtz, 2015). In another study, listeners accepted large deviations in how particular vowels were pronounced after only being exposed to small deviations in pronunciation (Witteman et al., 2010). In addition, after being presented with an accent in which stops were devoiced syllable-finally, listeners accepted mispronunciations where stops were devoiced syllable-initially, even if they were never devoiced in this position during the exposure period (Eisner et al., 2010). Finally, toddlers exposed to an accent where [a] was produced [æ] also accepted [ɛ] for [a] (White & Aslin, 2011). In all of these cases, adults and children alike seem to be allowing more flexibility in how a sound is produced, even if it does not align with what they previously heard in this accent.

Although the *Shift and Expand* hypothesis has not been directly studied in the accent adaptation literature, it is consistent with many of the findings that support the *Expand* hypothesis. These findings mostly provide evidence that categories expand during adaptation, without considering whether they also shift. In addition, recent work suggests listeners use this strategy in other related instances of category adaptation, including phonetic recalibration and selective adaptation, so perhaps they are also using it in accent adaptation (Kleinschmidt & Jaeger, 2015). When adults are played sounds that are ambiguous between two categories (e.g. /b/ and /d/) and then shown visual stimuli that disambiguate them, they adapt their categories in accordance with the visual stimuli (phonetic recalibration). On the other extreme, when people are repeatedly played prototypical instances of a category (e.g. /b/), they adjust their categories in a way that makes them less accepting of non-prototypical instances of this category (selective adaptation). A model in which categories are both shifted and expanded in response to speech is able to explain these results, suggesting that adaptation involves both shifting and expanding.

Most work on accent adaptation has found evidence for category expansion, but a key finding by Maye et al. (2008) stands as an exception. In their experiment, participants listened to passages of the Wizard of Oz, with all of the front vowels lowered (e.g. 'witch' was pronounced 'wetch'). After just twenty minutes of exposure, participants were more likely to accept items that were mispronounced with lowered vowels (e.g. 'wetch') as being words, but were no more likely to judge items mispronounced with raised vowels (e.g. 'weech') as words. From this result, Maye et al. (2008) intuit that people shift their categories, rather than expand them.

In this work, we model Maye et al.'s data to see whether their conclusion is warranted. We apply the ideal adaptor model from Kleinschmidt & Jaeger (2015) to compare three adaptation models: one in which categories are shifted, one in which they are expanded, and one in which they are both shifted and expanded. We compare these to a fourth model that directly learns the vowel mappings (e.g. that [i] is produced [ɪ] in this dialect). We show that the data from Maye et al. (2008), which have typically been used to argue against expansion hypotheses, can be captured by all four hypotheses. This suggests that the Maye et al. (2008) results may not provide strong evidence for the *Shift* hypothesis after all.

We begin with a description of the Maye et al. (2008) experiment, before describing our simulations.

Maye et al. (2008)

Maye et al. (2008) experimentally investigated how adults adapt their sound categories in response to accented speech, asking whether people shift or expand their vowel categories. The experiment was conducted in two sessions that were spaced a few days apart. In Session 1, participants listened to twenty minutes of the Wizard of Oz spoken in their accent by a synthesized male voice. The participants then took part in a lexical decision task, described below, in which they decided whether various test items they heard were words or not.

In Session 2, the same participants listened to the same passage; however, in this session, all of the front vowels were lowered. The vowel [i] was pronounced as [ɪ], [ɪ] was pronounced as [ɛ], [ɛ] was pronounced as [æ], [æ] was pronounced as [a] and [a] was unaltered, resulting in a merger between [æ] and [a]. None of the back vowels were lowered. After twenty minutes of exposure to this accented speech, in which participants heard fragments like 'the weckud wetch of the wast,' instead of 'the wicked witch of the west,' participants took part in the same lexical decision task from Session 1. Participants' responses between the two sessions were compared to see how exposure to accented speech affected their judgments.

Six types of test items were included in the lexical decision task. *Witch* items are words under the participants' accent, but are not words under the lowered accent ('witch' is a lowered version of 'weech', which is not a word in standard English). *Wetch* items are words in the lowered accent, but are not words in the standard American accent ('wetch' is a lowered version of 'witch'). *Weech* items are words in standard American English mispronounced with raised front vowels ('weech' is a raised version of 'witch'). The remaining three test item types did not contain any front vowels, so these would be pronounced identically under both accents. *Girl* items are words under both accents. *Loke* items are non-words under both accents, but are lowered versions of a word in standard English ('loke' is a lowered version of 'look'). *Tuke* items are non-words under both accents, but are raised versions of a word in standard English ('tuke' is a raised version of 'took'). Participants were presented with test items of these types - half of them occurred in the story and half of them did not - and were asked whether they were a word or not.

Maye et al. (2008) argue that if people are generally more flexible with how categories are produced (*Expand* hypothesis), then the participants should show an increase in endorsement rates for both *Weech* (raised) items and *Wetch* (lowered) items. On the other hand, they suggest that if listeners shift their categories in the direction of the accent they hear (*Shift* hypothesis), then participants should only show an increase in endorsement rates for *Wetch* items, not *Weech* items.

The results from their experiment are shown as part of Figure 2. Participants show an increase in endorsement rates for *Wetch* items, but not *Weech* items. Based on this finding, Maye et al. (2008) conclude that people are shifting their categories, not expanding them. In this paper, we model the data from this experiment and find that their results do not reliably support the *Shift* hypothesis over either expansion hypothesis (*Shift and Expand* or *Expand*). In the next section, we outline the modeling framework - the ideal adaptor framework - that we use to test these three hypotheses.

The Ideal Adaptor Model

We use the ideal adaptor model from Kleinschmidt & Jaeger (2015) to instantiate the *Shift*, *Expand*, and *Shift and Expand* hypotheses. We generalize the model to two dimensions, representing vowels as two-dimensional Gaussians. The vowels are defined by their first and second format values, which are continuous measures that are standardly used to distinguish between vowels. At any given point, the model has a set of categories, each of which is defined by a mean and covariance. With exposure to new speech, these categories change (the mean and covariance get updated) by taking a weighted average of the previous category information and the newly encountered data. We formalize shifting a category as changing its distribution's mean based on new evidence. We formalize expanding a category as updating the covariance of its distribution based on new evidence.

Broadly speaking, if the newly encountered data point lies near the category that generated it, then the category mean and covariance will be minimally changed, but if the newly encountered data point lies far from the category that generated it, then the mean and covariance will be more substantially changed to account for that data point. Crucially, the properties of this model are such that we can manipulate how much confidence we place on the previous mean and covariance, to manipulate how much the mean and covariance get updated. If we have high confidence in the previous mean (or covariance), then it will be less affected by new data, whereas if we have low confidence in the previous mean (or covariance), then the new data will affect the parameters more. We take advantage of this property to implement three of the hypotheses we would like to test. For the *Shift* model, we place high confidence in the previous covariance and low confidence in the previous mean. For the *Expand* model, we place low confidence in the previous covariance and high confidence in the previous

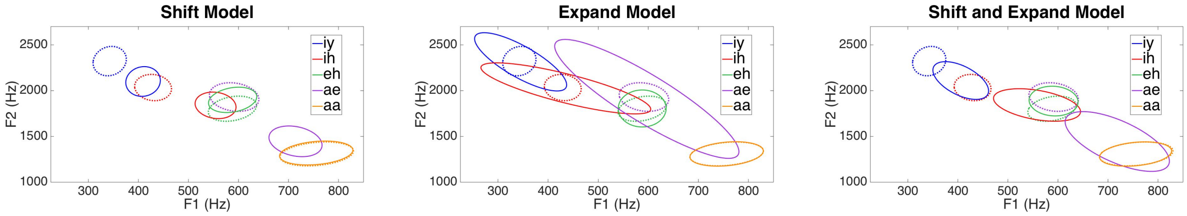

Figure 1: Original categories (dotted lines) and updated categories (solid lines) for each model.

mean. Finally, for the *Shift and Expand* model, we place low confidence in both the previous mean and covariance.

We assume a Normal Inverse Wishart prior over the category means and covariances based on speaker productions and assume that these means and covariances are updated as follows. The updated mean of a category after seeing N new acoustic values, μ_n, is the sum of the previous mean of the category, μ_0, weighted by the confidence in the previous mean, K, and the mean of the new acoustic values, $\bar{x}$, weighted by how many of them there are,

$$\mu_n = \frac{K}{K+N}\mu_0 + \frac{N}{K+N}\bar{x} \tag{1}$$

As confidence in the previous mean, K, increases, more weight will be placed on the previous mean than the mean of the new instances and the category mean will not change very much.

Similarly, the updated covariance of a category after seeing N new acoustic values, Σ_n, is the sum of the previous covariance, Σ_0, weighted by the confidence in the previous covariance, V, and the covariance contributed by the new acoustic values, weighted by the number of new values observed, N. The covariance contributed by the new acoustic values is a sum of the covariance of the new acoustic values, s^2, and a weighted term that accounts for the deviation of the mean of the new acoustic values from the previous mean,

$$\Sigma_n = \frac{V}{V+N}\Sigma_0 + \frac{N}{V+N}\left(s^2 + \frac{K}{K+N}(\mu_o - \bar{x})(\mu_o - \bar{x})^\top\right) \tag{2}$$

As confidence in the previous covariance, V, increases, the previous covariance will be weighted more heavily and the category covariance will not be updated as much. We use this ideal adaptor framework to implement our three adaptation models (see Kleinschmidt & Jaeger (2015) for more details of the generative model). The next section explains how we simulate the experiment performed by Maye et al. (2008).

Simulation of Accent Adaptation Experiment

To simulate the experiment, we first initialize the models with vowel categories based on standard American English, then update their categories based on the acoustic values participants would have heard in both experimental sessions, and, based on each model's updated categories, simulate the lexical decision task participants took part in. In the following sections, we outline each of these components.

Category Initialization

Each model was initialized with a category set that corresponded to a standard American English speaker. The standard accent categories were estimated based on data reported by Hillenbrand et al. (1995), in which participants produced vowels in a "h_d" context (e.g. 'head,' 'had,' 'heed'). Because the speech in the experiment was spoken by a male voice, we limited our estimates to vowels produced by male speakers. For each vowel type, there were 45 productions, each produced by a different speaker. For each vowel category, we took the 45 productions in F1-F2 space and calculated the mean and covariance of these 45 data points, which we used as the mean and covariance of the Gaussian distribution representing that category. These were the categories the models started with (the categories corresponding to the standard accent), shown in dotted lines in Figure 1.

Familiarization Data

We approximated what the participants in the experiment heard in both of the experimental sessions. Based on average reading times, the participants would have heard about 2500 words spoken during the 20-minute story excerpt (Olive et al., 1993). To get the actual items, we automatically transcribed all of the words in the Wizard of Oz, using the CMU Pronunciation Dictionary (Weide, 1998), and then manually verified that the results were accurate. We sampled 2500 words in proportion to how frequently they occur in the story. The same words were presented to the model in Session 1 and Session 2. For each vowel in the presented words, a particular formant-value pair was sampled. The distribution from which the vowel token was sampled depended on the session and the frontness of the vowel. For both front and back vowels in Session 1, a token was sampled from the distribution corresponding to that vowel in standard American English. For Session 2, front vowels were sampled from the standard American English distribution corresponding to the lowered vowel (for example, [i] tokens were sampled from the [ɪ] category) and for back vowels, the same formant values were repeated from Session 1. Throughout the experiment, the test item vowels were sampled from the same underlying distributions, which represent the (unchanged) sound categories of the standard American dialect. The listeners' categories were updated based on this exposure.

Category Adaptation

For the *Shift*, *Expand*, and *Shift and Expand* hypotheses, we update the models' categories according to (1) and (2). We also implement a fourth model, in which we update each category by setting its mean and covariance to the mean and covariance of its lowered version (*Vowel Mapping* model). For example, the updated [i] category will have the mean and covariance of

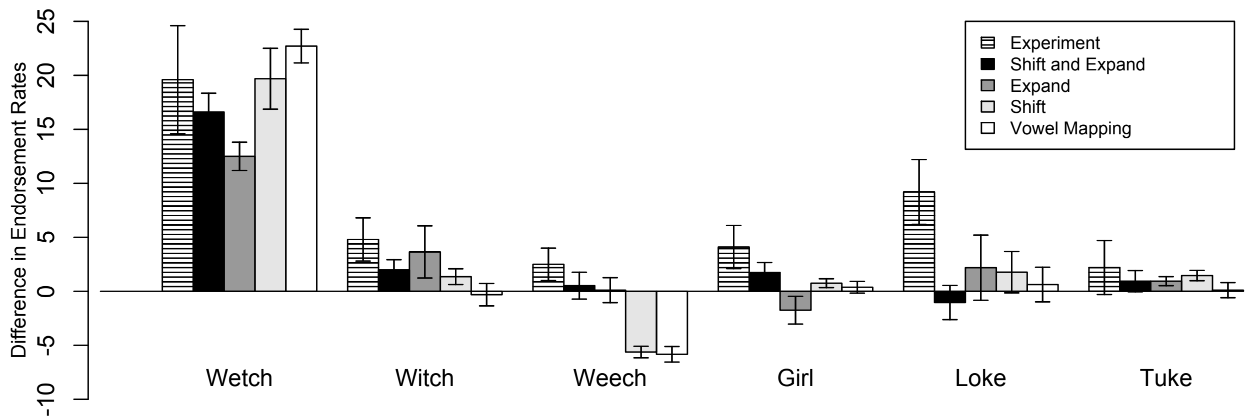

Figure 2: Difference in endorsement rates between Session 1 and Session 2 from Maye et al. (2008) and models.

the original [ɪ] category, the updated [ɪ] category will have the mean and covariance of the original [ɛ] category, and so forth.

There is some evidence reviewed in Kleinschmidt & Jaeger (2015) that listeners build different sets of categories for different groups of speakers. For example, people could learn that adult men speak differently than toddlers and instead of representing speech by both groups as identical, they could build two sets of categories to capture their speech differences. They might have one /i/ category corresponding to adult males and another /i/ category corresponding to toddlers.

Based on these results, when the model encounters the accented speech from Session 2, a new set of categories is built to represent the accented speech. These categories are initialized where the previous categories were, but only these newly-formed categories are updated in response to this new type of speech. At the end of the experiment, the model has one set of categories for speech spoken in a standard accent and another set for speech spoken in a lowered accent.[1]

As discussed above, we manipulate the speed of the adaptation by choosing particular values of K and V (confidence in previous mean and variance). Based on parameter settings found to capture experimental data in Kleinschmidt & Jaeger (2015), we set both K and V to 100 for the *Shift and Expand* model, we set K to 100, but V to 10000 for the *Shift* model, and we set K to 10000 and V to 100 for the *Expand* model.

Lexical Decision Task

After the categories are updated in Session 1 and Session 2, a lexical decision task is simulated. We adapt the model of lexical decision developed by Norris (2006) for written word recognition tasks to the problem of spoken word recognition tasks. The goal in a lexical decision task is to determine whether a test item is a word or not, not to identify the particular word. As a result, inference consists of inferring the word status (w) of a test item given its form. To do so, we apply Bayes' rule to calculate the relative probability that a test item is a word, given the acoustic input, x,

$$P(w = word\,|\,x) = \frac{P(x\,|\,w = word)P(w = word)}{\sum_w P(x\,|\,w)P(w)} \tag{3}$$

That is, the posterior probability that the test item is a word is proportional to the product of the probability that the particular acoustic value in question was generated from a category that would make the test item a word (likelihood) and the prior probability that a test item is a word.

Because there are two category sets at the end of the Session 2 (one set corresponding to unaccented speech and another corresponding to the accented speech), we need to sum over both of these category sets (c=standard and c=lowered),

$$P(w = word\,|\,x) = \frac{\sum_c P(x\,|\,w = word, c)P(w = word)P(c)}{\sum_{w,c} P(x\,|\,w,c)P(w)P(c)} \tag{4}$$

This is because we need to incorporate all possible ways that the item could have been generated. For example, 'wetch' could be an unaccented pronunciation of the non-word 'wetch' or it could be an accented pronunciation of the word 'witch' and both of these possibilities need to be considered.

In order to calculate the probability in (4), we need the values of $P(x\,|\,w = word, c)$, $P(w = word)$, and $P(c = standard)$. We consider each of these in turn, beginning with $P(x\,|\,w = word, c)$. In theory, we need to sum over all possible words, L, that the test item could be an instance of to get this probability. For example, if we want to calculate the posterior probability of 'wetch' being a word, we need to sum the possibility that this was a mispronunciation of the word 'watch' under standard English, a mispronunciation of 'witch' under standard English, a mispronunciation of 'watch' under accented English, etc. Because most of the elements in this sum are close to zero, we make the simplifying assumption that participants are only considering the possibility that one specific word or one specific non-word generated the input (5). All of the test items are either a word, or based on a word (e.g. 'wetch' is a lowered version of 'witch'), so we make use of the most relevant word, l^*, and non-word in our calculations. For example, if participants hear 'witch,' they infer whether this came from

[1] We also ran a version of each model where we directly updated the initial categories, without building a new set of category to represent the accented speech. These fail to fit the data because they predict a substantial decrease in the endorsement of *Witch* items, which is not observed in the experiment.

Model	P(c = standard)	P(w = word)
Shift	0.75	0.9
Expand	0.9	0.9
Shift and Expand	0.9	0.9
Vowel Mapping	0.95	0.9

Table 1: Parameter setting that produced results that best matched Maye et al. (2008) for each model.

Model	Log-Likelihood
Shift	-394.82
Expand	-388.72
Shift and Expand	-388.04
Vowel Mapping	-395.97

Table 2: Log-likelihood of models given Maye et al. data. Less negative values indicate a better match with the data.

a word or a non-word, by considering the probability that it was generated from the word 'witch' or the non-word 'weech.'

$$P(x|w = word, c) = \sum_{l \in L} P(x|l, c) \approx P(x|l^*, c) \quad (5)$$

To get the probability of the phonetic form of the word given the particular lexical item and category set, $P(x|l^*, c)$, we take the vowel token in the word in F1-F2 space and calculate the value of the Gaussian probability density function of the relevant vowel category at this point. This probability is highest near the participant's category means.

The remaining two terms are free parameters: one corresponds to the probability of a test item being a word, $p(w = word)$, and the other corresponds to the probability of using the standard English set of categories, $p(c = standard)$. To deal with these two free parameters, each version of the model was run on all combinations of these two parameters, ranging from 0.05 to 0.95 by increments of 0.05, for a total of 200 combinations. Because there is quite a bit of randomness involved throughout the process, the model was run twenty times on each parameter combination and the results were averaged across these twenty runs.

Putting this into (4), we get the relative probability of the input, x, being a word versus not. The model classifies an item as a word or non-word by sampling from its posterior distribution, $P(w|x)$.

To summarize, we update the model's categories in one of three ways: by shifting them, expanding them, or shifting and expanding them. In addition, we implemented a fourth model that sets categories to their lowered version. We compare the four models' results to human performance to see which can capture the experimental findings.

Simulation Results

The front categories before and after they are updated are shown in Figure 1. The next two sections compare the models' performance to the Maye et al. (2008) results.

Difference in Endorsement Rates

Figure 2 shows the difference in endorsement rates between Session 1 and Session 2 from the experiment and for each model. The figure shows each model's results from the parameter settings that had the best match with the original data (shown in Table 1). This was determined by calculating the log-likelihood of the model (i.e. the probability of the differences in endorsement rates observed in the experiment given the model). These results are shown in Table 2 - less negative log-likelihoods indicate better matches. Because all four models have the same number of free parameters, model comparison methods like the Bayesian information criterion (BIC) reduce to simply comparing the model log-likelihoods.

Overall, comparing the log-likelihoods, all three adaptation hypotheses outperform the *Vowel Mapping* model. Qualitatively, all four models are able to capture the increase in endorsement rates for *Wetch* items between sessions, though the *Expand* model underestimates this increase. The *Expand* model is, however, also able to capture the slight increase in endorsement rates observed for *Weech* items.

The *Shift* and *Vowel Mapping* models have the worst log-likelihoods, mainly because they both incorrectly predict a slight decrease in endorsement rates for *Witch* items between the two sessions. This happens because the free parameter settings that result in an increase in endorsement rates for *Wetch* items (i.e. high reliance on the new category set) also result in a decrease in endorsement rates for *Witch* items. Although this tension exists for all of the models, the *Expand* and *Shift and Expand* models overcome this because of their high-variance updated categories that continue to overlap with the original category. The *Shift* and *Vowel Mapping* models share the property that their updated categories do not overlap with the original categories, which explains their similar predictions.

It is worth noting that none of these models capture the *Loke, Tuke,* or *Girl* results very well. The participants seem to be generalizing the lowering of vowels to back vowels to a certain extent, which our models cannot capture because back vowels are left completely unaltered.

Overall, despite the conclusions that have been drawn from these results, all three adaptation processes (*Shift*, *Expand*, and *Shift and Expand*) can explain them to approximately the same degree, with the adaptation processes involving category expansion performing slightly better.

Absolute Endorsement Rates

Figure 3 shows the absolute endorsement rates for each of the test item types before and after Session 2 for the participants and the *Shift and Expand* model. The model underestimates people's absolute endorsement rates for all of the items, except *Girl* and *Witch* items. In particular, before they are even exposed to the new speech, participants accept *Wetch* items as words 39% of the time, whereas the model barely reaches this endorsement rate after updating its categories with the unfamiliar accent. Similarly, the model accepts *Weech* items around 35% of the time, but participants accept them as words between 60-70% of the time. Participants accept *Loke* and *Tuke* items about 30% of the time, but the model accepts them

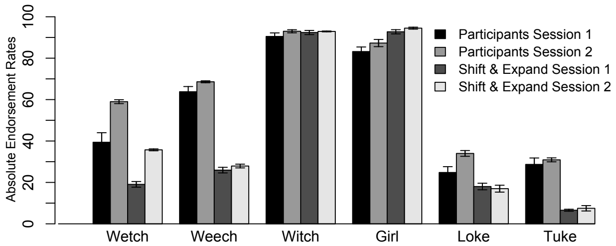

Figure 3: Absolute endorsement percentages by participants in Maye et al. (2008) and *Shift and Expand* model.

far less often. On the other hand, the model predicts that participants accept *Girl* item types over 90% of the time, despite the fact that participants are only accepting these words around 80% of the time. Therefore, there are aspects of the experimental data that the model does not seem to be capturing.

One possible explanation for why we observe this difference between the participants and our models is that unlike our models, people have had exposure to many different types of accents before. If they have heard accents that affect vowels in similar ways as the accent in this experiment, then perhaps they already have a representation for this type of speech and accept it more willingly. The models have only had exposure to a standard American accent. Incorporating some knowledge of other accents would be likely to increase the models' acceptance of these test item types.

Discussion

This paper explores how people adapt their sound categories in response to novel accents. We simulated the Maye et al. (2008) experiment using four different models, each of which adapted categories differently. In one version, categories were adapted by being shifted, but not expanded (*Shift* hypothesis). In a second version, categories were expanded, but not shifted (*Expand* hypothesis). In the third model, categories were shifted and expanded (*Shift and Expand* hypothesis). The final *Vowel Mapping* model directly learned the category mapping between accents. All models captured the data reasonably well, though the expansion models seemed to fare slightly better than the *Shift* and *Vowel Mapping* models. Crucially, this was the case on data that have been used to argue that people use the *Shift* strategy, rather than the *Expand* strategy.

How is *Expansion* able to capture the data? Maye et al. (2008) argue that only a *Shift* strategy can explain why participants show an increase in endorsement rates for *Wetch* items, but not *Weech* items, yet we found that expansion hypotheses were able to capture this asymmetry. Their assumption of symmetry is true if the affected vowels are all equidistantly spaced along a line. In any space where this is not true, such as F1-F2 space, their argument does not hold. On top of that, their test item groups differed in the number of items involving each vowel pair. For example, there were fifteen *Weech* items involving [ɪ] raising to [i], but only six *Wetch* items involving [i] lowering to [ɪ]. Because some vowels are closer together than others, such differences could produce endorsement rate asymmetries between different test item types that are not reflective of listeners' adaptation strategy. Contrary to their assumption, an *Expand* model can lead to asymmetric responses in lexical decision tasks, as they observed in their data.

We only tested one instantiation of each of the *Shift*, *Expand*, and *Shift and Expand* models; therefore, the conclusions we can draw about which of these models best capture the experimental findings are limited. There may, for example, be other instantiations of the *Shift* hypothesis (i.e. with different mean and covariance confidence parameters) that outperform the models discussed here. Overall, however, all three adaptation models are able to capture, to a better or worse degree, the main qualitative findings from Maye et al. (2008).

Most work on accented speech perception has suggested that category expansion is a component of accent adaptation. A key exception comes from Maye et al. (2008), who argue that people shift their categories, without expanding them. Although their experimental results may intuitively support a *Shift* strategy, building a model of the results shows that this argument is not conclusive. In fact, the *Shift & Expand* and *Expand* strategies explain these findings equally well to, or potentially better than the *Shift* hypothesis. Taken together with the rest of the accent adaptation literature, this work suggests that people may be relying on expansion strategies in accent adaptation.

Acknowledgments We thank the reviewers, the UMD Probabilistic Modeling group, and the audience at NECPhon 9 for helpful feedback. This work was supported by NSF grant IIS-1421695.

References

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented english. *The Journal of the Acoustical Society of America*, *116*(6), 3647-3658.

Eisner, F., Weber, A., & Melinger, A. (2010). Generalization of learning in pre-lexical adjustments to word-final devoicing. *The Journal of the Acoustical Society of America*, *128*(4), 2323–2323.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099–3111.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562.

Norris, D. (2006). The bayesian reader: explaining word recognition as an optimal bayesian decision process. *Psychological review*, *113*(2), 327.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of american english speech: a dynamic approach*. Springer Science & Business Media.

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, *15*(6), 732–738.

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*. Doctoral dissertation, The Ohio State University.

Weide, R. (1998). *The cmu pronunciation dictionary, release 0.6*. Carnegie Mellon University.

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384.

Witteman, M. J., Weber, A., & McQueen, J. M. (2010). Rapid and long-lasting adaptation to foreign-accented speech. *The Journal of the Acoustical Society of America*, *128*(4), 2486–2486.