Failure to replicate talker-specific syntactic adaptation

Linda Liu[1], **Zachary Burchill**[1], **Michael K. Tanenhaus**[1,2], **T. Florian Jaeger**[1,2,3]
{lliu47, zburchil, mtanenha, fjaeger} @ ur.rochester.edu
Department of Brain and Cognitive Sciences[1], Department of Linguistics[2],
Department of Computer Science[3],
University of Rochester
Rochester, NY 14627 USA

Abstract

Sentence understanding is affected by recent experience. An important open question is whether this reflects adaptation to the statistics of the input. Support for this hypothesis comes from the recent finding that listeners can simultaneously learn and maintain the syntactic statistics of multiple talkers (Kamide, 2012). We attempt—and fail—to replicate this finding. This calls into questions whether recency effects in sentence processing originate in the same adaptive mechanisms operating during speech perception (for which talker-specific adaptation is well-established).

Keywords: sentence processing; attachment ambiguity; priming; adaptation; talker specificity

Introduction

Talkers differ in how they realize the same sound categories and words. This inter-talker variability is known to be one of the biggest challenges to speech perception, and has been investigated by a large body of research. This work has identified adaptation as a central mechanism by which listeners overcome inter-talker variability (e.g., Bradlow & Bent, 2008; Kraljic & Samuel, 2007). With exposure to a novel talker, listeners seem to be able to adapt their category boundaries to the statistics of the current input (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015). This adaptation can persist over time (Eisner & McQueen, 2006), and listeners can maintain separate adaptations for different talkers (Kraljic & Samuel, 2007; Trude & Brown-Schmidt, 2012). Findings like these suggest that listeners continuously infer and store talker-specific information about the realization of phonological categories (for review, see Kleinschmidt & Jaeger, 2015).

Here we focus on adaptation beyond speech perception, and the extent to which it exhibits properties similar to those demonstrated for speech perception. It has long been recognized that talkers also vary in, for example, their lexical and syntactic preferences, and that these differences can impede processing (for a recent review, see Fine, Jaeger, Farmer, & Qian, 2013). For example, the efficiency of expectation-based processing (i.e., utilizing expectations based on previous input to facilitate integration of bottom-up input) depends on the degree to which comprehenders' expectations—based on previously experienced input—match the statistics of the present input. Thus, if talkers differ in their preferences, and thus the statistics of the input they provide, this should impede processing. On the flip side, expectations that match the statistics of the current input will facilitate processing.

This and related considerations have motivated work on adaptation beyond speech perception, including lexical (e.g., Creel, Aslin, & Tanenhaus, 2008), prosodic (e.g., Kurumada, Brown, & Tanenhaus, 2012), and—most relevant to the present purpose—syntactic adaptation (e.g., Fine et al., 2013). For syntactic processing, there is now some evidence that experience with a novel environment (e.g., an experiment) can lead to longer-lasting changes in processing, persisting for at least several days (Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Further, paralleling studies on adaptation in speech perception, a recent study found that listeners can adapt to multiple talkers simultaneously (Kamide, 2012). As one of the few studies to demonstrate talker-specificity in syntactic adaptation, Kamide's study constitutes an important contribution to our current understanding of whether syntactic adaptation is best understood in terms 'dumb' priming mechanisms (Tooley, 2009) or 'smarter' learning mechanisms (Fine et al., 2013). This result has received a fair amount of attention (with about 10 citations/year); however, as we will discuss later, there are some reasons to interpret the finding with caution.

Here we report efforts to replicate Kamide (2012). We begin by summarizing the original study. Then we introduce our own paradigm, which closely follows this study (we gratefully acknowledge Yuki Kamide's generous willingness to discuss details of her design; errors remain our own). We note here that we made a few changes to the procedure and stimuli of Kamide's paradigm, which we introduce and motivate as we go through our experiment.

Overview of Kamide (2012)

Kamide (2012) employed a look-and-listen visual world eye-tracking paradigm to investigate changes in listeners' syntactic expectations for specific talkers. Participants saw visual scenes like that depicted in Figure 1.

These scenes were paired with sentences containing a syntactic attachment ambiguity. For example, in the sentence "The master of the dog who will bury the treasure is quite old now.", the relative clause, "who will bury the treasure," can describe "the master" or "the dog". These two interpretations correspond to different syntactic parses. The former constitutes *high attachment*, the latter *low attachment*.

This attachment ambiguity is assumed to be resolved at "treasure" since MASTER is more likely to be the agent of the BURYING-TREASURE event. Kamide's analyses focuses on the temporarily ambiguous stretch between the introduction

Figure 1: Example experimental scene paired with example (high attachment) sentence, both from Kamide (2012). Underlining marks the syntactically ambiguous sentence region, during which saccades to the target (treasure) and competitor (bone) objects are taken to indicate high vs. low attachment, respectively. (Target and competitor objects are not labeled in actual stimuli shown to participants).

of the relative clause verb and the resolution of the ambiguity ("will bury"). During this stretch, eye-movements to the master or dog can be seen as reflecting expectations that the relative clause will attach high (or low). Specifically, listeners should show more anticipatory eye-movements to referents that are plausible objects for the expected agent. In Figure 1, if high (low) attachment is expected, listeners should show anticipatory eye-movements to the treasure (bone).

This property theoretically makes it possible to investigate *changes* in syntactic expectations based on recent exposure, including changes in expectations *for specific talkers*. Specifically, Kamide exposed participants to sentences from three different talkers. One talker always used high attachment sentences, one talker always used low attachment, and the final talker used high and low attachment equally often. The question then is whether participants' eye-movements during the temporarily ambiguous stretch begin to reflect talker-specific syntactic preferences. If so, participants should be more likely to exhibit anticipatory saccades toward the target object for the two talkers that consistently attached either low or high (cued talkers), compared to the talker who used high and low attachment equally often (uncued talker). This advantage should emerge with increasing exposure throughout the experiment.

Kamide tested this prediction by analyzing the number of trials in which at least one saccade was launched to each object (target vs. competitor) for the cued vs. uncued talkers during each phase of the experiment (beginning [first two blocks] vs. end [last two blocks]). Supporting the hypothesis of talker-specific syntactic adaptation, Kamide identified a statistically significant three-way interaction between object x cue x phase: at the end of the experiment (compared to the start), participants made more saccades to the target for the cued talkers, compared to the uncued talkers.

However, there are reasons to interpret this result with caution. First, there were relatively few trials with fixations to either the target (12.1%, averaged across the beginning and end of the experiment) or the competitor (11.8%). Second, the looks to the target and the competitor actually *de*creased over the course of the experiment (from 15.0% to 9.2% for the target and 15.6% to 8.0% for the competitor). Both of these effects result in a small signal to detect potential effects, which we try to address by using a task-based paradigm.

Experiment

Methods

Participants We recruited 24 participants at the University of Rochester (c.f. 48 participants in the original study). Although we had originally intended to run 48 participants, a power simulation (reported below) following 24 participants suggested we would be unlikely to detect an effect, even with the additional participants. All participants were native speakers of English with normal vision. The experiment took about 1 hour and participants were paid $10.

Visual Stimuli We used the visual scenes (20 experimental and 40 filler) from Kamide (2012). Each experimental scene consisted of six images arranged against a colored background (e.g., Fig. 1). Two images corresponded to the target and competitor agents of the sentence, and four images corresponded to the potential objects for the sentence.

Sentences We used the original sentences from Kamide (2012), with minor vocabulary adjustments to accommodate American listeners (e.g., motorcycle instead of motorbike). Each scene was paired with four training sentences (A) and four test sentences (B), which involved novel objects referenced in the training sentences. Sentences 1 and 3 are sentences with high attachment interpretations (i.e., in 1A, the master is the agent of the main verb *bury*). Sentences 2 and 4 are sentences with low attachment interpretations (i.e., in 2A, the dog is the agent of the main verb *bury*).

1. The master of the dog who will *(A) bury the treasure / (B) drink the brandy* is quite old now.
2. The master of the dog who will *(A) bury the bone / (B) drink the water* is quite old now.
3. The dog of the pirate who will *(A) bury the bone / (B) drink the water* is quite old now.
4. The dog of the pirate who will *(A) bury the treasure / (B) drink the brandy* is quite old now.

Additionally, we used the 40 original filler sentences which did not contain a syntactic ambiguity, and instead contained a relative clause containing an unambiguous single noun phrase antecedent (e.g. "The woman who will lift the pet carrier has never had a dog.").

Auditory Stimuli American listeners might have difficulty distinguishing between the varieties of UK English used in the original study (conducted in Scotland). We thus recorded new materials from talkers selected to be clearly distinguishable by American listeners. We recorded one British-accented male (BM; average sentence duration = 3,668 ms, SD = 485 ms), one British-accented female (BF; 3,944 ms, SD = 430 ms), and one Indian-accented Male (IM; 4,358 ms, SD = 516 ms). The BM and BF talkers served as the cued talkers, consistently producing either only high attachment sentences or only low attachment sentences. Between participants, we counterbalanced the assignment of high vs. low attachment to talkers, so that each talker served as the high attachment talker for half of participants. Following Kamide, one talker (Talker IM) always served as uncued talker, producing equal amounts of high and low attachment sentences.

All audio recordings were scaled to an average intensity of 70 dB. Following Kamide (2012), the shared portion of each low attachment sentence (up to the relative clause; i.e. "the master of the dog") was then spliced into the corresponding high attachment sentence to ensure that any effects found would not be due to early prosodic cues; i.e. both high and low attachment sentences shared the same low attachment prosody.

Presentation Lists Following Kamide (2012), we created four presentation lists, crossing whether Talker BM and Talker BF served as the high and low attachment talkers, and which items were used in the cued vs. uncued talkers. Each participant was tested on one of these lists.

Each list was divided into six blocks, each consisting of 30 trials (20 experimental, 10 filler). Each experimental scene was presented once per block, with a different sentence each time (i.e. in the first four blocks, sentences 1-4A were heard, and in the final two blocks, either sentences 1-2B or 3-4B were heard). During each block, the uncued talker (IM) produced *both* five low and five high attachment sentences, and the cued talkers (BM and BF) each produced *either* five low or five high attachment sentences, along with five filler sentences. Sentence order within the lists was pseudorandomized following the criteria described in Kamide (2012).

Procedure Our procedure closely followed that of Kamide (2012) with the exception that we used a task-based clicking paradigm (described below) as compared to a look-and-listen paradigm. This decision was made because the results of the original study suggest that there was very little signal to detect potential effects. First, in the original study, the percentage of trials in which a subject makes a saccade toward the target or the competitor in the critical window was small (fewer than 24%). Second, participants show a (highly significant) decrease in saccades to both target and competitor over the course of the experiment: 15.0% to 9.2% for the target object and 15.6% to 8.0% for the competitor object.

In a comparison between a look-and-listen paradigm and an explicit task-based paradigm, Altmann and Kamide (1999) find that although results between the two paradigms were qualitatively similar, the task-based paradigm had less-delayed anticipatory eye movements, and subjects were almost twice as likely to fixate relevant referents (see Salverda & Tanenhaus, in press).

Expecting more looks to the target and competitor (as task-relevant objects), we incorporate the task of clicking on the target into Kamide (2012). This carries the additional benefit of allowing us to test whether participants are paying attention during the experiment (see Results).

Participants were seated in front of a 19 inch computer screen. Eye movements were monitored using an Eyelink 1000, sampling at 500 Hz from the left eye. Recalibration was performed at the beginning of each new block (every 30 trials), and drift correction was performed every five trials.

At the start of each trial, participants saw a task prompt against a white background on the screen, e.g. "Click on what will be buried." Crucially, the object that should be clicked on varied between high and low attachment sentences (i.e. treasure or bone). After the participant clicked to verify that they had read the prompt, a fixation cross appeared in the center of the screen. After the participant clicked on the fixation cross, the visual scene appeared on the screen. The sentence played over loudspeakers after the scene preview period of 1000 ms elapsed. The trial ended after participants clicked in the scene after the audio file ended.

Results

We begin by assessing how participants engaged in the experimental task. First, we test whether participants performed the task as intended. Second, we assess whether the task-based paradigm caused participants to have more saccades to the task-relevant referents during the critical time window, thereby increasing our power to detect an effect. Third, we test whether the number of trials containing saccades to target, compared to competitor, objects decreases over the course of the experiment (as reported in the original study).

All analyses presented below include the maximal random effects justified by the design (i.e. by-participant and by-item intercepts and slopes for cue, phase, and their interaction). We treat sentences sharing the same prefix up to the relative clause as one item (e.g., 1A-B and 2A-B form an item), such that cue, phase, and their interaction varied within items. We continue to use the term **target object** to refer to the object that is consistent with the intended attachment interpretation (e.g., the treasure in 1A), and the term **competitor object** to refer to the object consistent with the *un*intended attachment (e.g., the bone in 1A).

Task performance: Click responses We find that participants overwhelmingly (98.3% of all trials) clicked on the correct target object (e.g., treasure in Fig. 1). This suggests that participants understood and successfully performed the experimental task. One scene showed a high incorrect response rate, with over 40% of subjects (10 of 24) clicking on the wrong object. This was likely due to the use of target label (*hat*) that is also a valid label for the competitor (*cap*). Trials involving this scene, and all trials with incorrect clicks, are excluded from subsequent analyses.

Task performance: Saccades to task-relevant referents Next, we analyze the number of trials containing looks to task-relevant objects (i.e., either of targets and competitors) and the changes in this number over the course of the experiment. We analyze the same time-window as in the original study: from the start of the onset of the verb of the relative clause to the onset of the object of the relative clause (e.g. "bury the" in Fig. 1). The mean duration for this time-window is 371 ms (cued; SD = 66 ms) and 430 ms (uncued; SD = 82 ms). These durations are somewhat different from, but similar to, the corresponding time-windows in the original study (cued: 528 ms, uncued: 365 ms).

Compared to Kamide, we find nearly twice as many trials containing saccades to the target (24.1% here vs. 12.1% in the original) or competitor objects (22.6% here vs. 11.8% in the original), averaged over the first and last two blocks. Like in the original study, we find that the overall probability of saccades to task-relevant objects decreases over the course of the experiment. However, this decrease was less dramatic than in the original study. In the first two blocks, 46.8% of all trials contained saccades to task-relevant referents and only decreased to 44.6% of all trials in the last two blocks. In fact, the number of trials in which there were saccades to the *target* object actually increased slightly, going from 22.2% in the first two blocks to 25.9% in the last two. This constitutes an *increase* of 0.2 in log-odds for the target object (compared to a decrease of 0.56 in the original study) and a decrease of 0.28 in log-odds for the competitor object (compared to a decrease of 0.75 in the original).

In summary, the task-based paradigm increases saccades to the task-relevant referents, as intended. This means that our estimates of the *relative* proportions of eye-movements to target vs. competitor objects—the dependent variable for the main analyses presented below—are based on more data. This should result in more reliable statistical signal for the main analysis (but see our discussion for caveats).

Main analysis: Anticipatory saccades during original time window We first analyze saccades during the time window employed in the original study. Failing to find evidence for talker-specific learning, we then present additional post-hoc analyses. Specifically, we extend our analyses to other time windows, and to individual cued talkers. All our analyses analyze the proportion (of eye-tracking samples with) fixations to the target vs. competitor objects during the time window. This differs from Kamide's, who analyzed the *number of trials containing at least one saccade* to the target vs. competitor during the time window. The two measures yield the same conclusions for the present data (see below). We chose our analysis approach, because Kamide's measure resulted in a model that converged only under a drastically reduced maximal random effects structure (by-participant random intercepts and slopes, and by-item random intercepts). We also note that our approach is known to be anti-conservative (because of auto-correlations between eye-tracking samples), but not known to be underpowered.

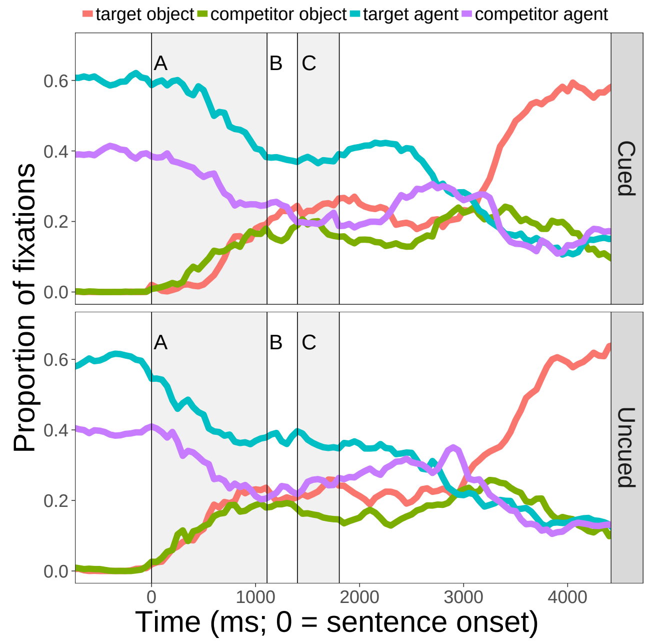

Figure 2: Proportion of fixations over the course of a trial. Colors show the different referents in the scene. Panels show the cued and uncued talkers. Regions A, B, and C correspond to the time windows in which participants heard the equivalents to "The master of the dog", "who will", "bury the", respectively. Kamide (2012)'s analysis was restricted to region C, as was our original analysis. We also present post-hoc analyses of regions A and B+C. For this plot, we normalized times within each region (essentially aligning all trials) before averaging across trials and subject.

The full time course for fixations over trials is shown in Fig. 2. Following Kamide (2012), we compare the data collected during the first two blocks (beginning) to the data collected during the final two blocks (end) during the time window between the start of the verb of the relative clause to the noun (e.g. "bury the").

We perform mixed logistic regression predicting fixations to target versus competitor from cue (sum-coded: cued = 1 vs. uncued = -1) and phase (sum-coded: end = 1 vs. beginning = -1). If participants are able to learn the syntactic preferences of specific talkers, then we expect more looks to the target object during the end of the experiment compared to the beginning. We would only expect this for the cued talkers, in which the talkers produced high or low attachment sentences consistently, and not for the uncued talker, in which the talker produced a mixture of both high and low attachment sentences. The critical result to replicate is the interaction between cue and phase of the experiment.

None of the factors reached significance (*ps* >0.14). The analysis is summarized in Table 1 (penultimate column).

Critically, we did not find a significant interaction between cue and phase ($p = 0.8$). Using the reduced model based on Kamide's measure, where the critical measure is a 3-way interaction between object x cue x phase, confirmed this pattern. Specifically, Kamide's measure returned a significant two-way interaction between object and cue in the unexpected direction ($\hat{\beta}$ = -0.08, p <0.05). The critical 3-way interaction trended in the same direction as in the original study, but did not reach significance ($p = 0.11$).

For the remaining analyses we only present analyses over proportions of fixations. Using Kamide's measure led to models with reduced random effect structures, but did not change the results.

Post-hoc analyses over additional time windows We entertain two possibilities for the observed null effect. First, it is possible that the introduction of a task-based paradigm affected the time course of eye-movements (see e.g., Altmann & Kamide, 1999). Specifically, in our paradigm listeners saw the verb prior to hearing the sentence, which may have allowed them trigger anticipatory eye-movements even earlier, before the onset of the RC verb. We therefore conduct post-hoc analyses over two additional time windows (see Figure 2):

1. Start of the sentence to start of the relative clause (SentenceStart to RCStart; e.g. "The dog of the pirate")
2. Start of the relative clause to start of the relative clause noun (RCStart to RCNounStart; e.g. "who will bury the")
3. Start of the relative clause verb to start of the relative clause noun (RCVerbStart to RCNounStart; e.g. "bury the")

A second possibility is that participants look towards the *agent*, rather than object, of the verb. We thus also analyze looks to the target versus competitor agents. For example, in the low attachment sentence (1A), the **target agent** is the dog, and the **competitor agent** is the master. We do so also for the original time window analyzed in the previous section.

This results in 6 analyses (3 time windows by 2 ways to operationalize target vs. competitor), of which one is the original analysis. The results from all analyses are presented in Table 1. None of the critical effects reach significance in any of the analyses. However, for the agents, we identified marginal interactions of cue and phase in two of the time windows (Table 1); however, both were numerically in the opposite of the expected direction. Following Kamide, we also performed each of these analyses for trials containing saccades to target or competitor (with a reducted random effects structure). None of the additional analyses reached significance in the predicted direction (predicted direction: $p > 0.7$), though the agents analysis for region C was significant in the unpredicted direction ($p < 0.05$).

We did not find a reliable effect of talker-specific syntactic adaptation between the cued and uncued talkers, even when considering post-hoc analyses over additional time-windows. All remaining analyses therefore use the same time windows as the original study.

So far, following Kamide (2012), all our analyses compared eye-movements for cued vs. uncued talkers. Such analyses group the two cued talkers together, ignoring potential differences in listeners' expectations for a male and a female (British English) talker. We conducted several follow-up analyses, comparing eye-movements in response to these two cued talkers. We found no evidence that participants were biased toward expecting more high (or low) attachment for either talker. Nor did we find any evidence that participants adapted more for one talker over the other.

General Discussion

In contrast to the growing literature on talker-specific adaptation in speech perception, one of the few studies that has directly addressed talker-specific syntactic adaptation is (Kamide, 2012). Here we were unable to replicate the original study's findings. We are aware of one other ongoing effort within the same paradigm as the original study, albeit with different stimuli (ongoing work by Ryskin, Fine, & Brown-Schmidt). Ryskin and colleagues do not find evidence for talker-specific syntactic adaptation over the original time window, while finding weak evidence over a larger time window.

One possibility why the present experiment failed to find talker-specific syntactic adaptation is that it had less power than Kamide's original study: we used fewer participants (24 here vs. 48 in the original study). On the other hand, presumably because of the use of a task-based paradigm in the present experiment, the proportion of trials containing a look to either the target or competitor was almost twice as high in the present study than in the original study (42% in the present study vs. <24% in the original study), which should increase the power. Power analyses suggested that power in our experiment was slightly lower than, but comparable to, Kamide's. For example, even when assuming an increase of 0.6 in log-odds in looks to the target in the cued vs. uncued condition from beginning to end of the experiment (constituting a 9-fold increase in effect size), the present approach (24 participants, 40% looks to task-relevant objects) yields power of 69.4%, compared to 82.5% for Kamide's study (48 participants, 20% looks to task-relevant objects).

Another possibility is that limited familiarity with the British- and Indian-accented English could have affected participants' ability to understand our talkers. If participants were unable to parse our talkers' sentences, this would make it impossible to learn talker-specific syntactic statistics. We note, however, that participants almost always clicked on the correct object (98.3% of all trials), indicating that they were able to successfully understand talkers. Additionally, the three talkers were easily distinguishable, which would be expected to support the learning of talker-specific statistics.

Additionally, it might be that our task encouraged participants to listen for the correct object without processing the syntactic structure. Participants were asked to click on the object of the verb; the verb was always provided at the very end of the sentence, and participants were not allowed to click un-

Table 1: Output of mixed effects logistic regression comparing fixations to target versus competitor objects and agents (positive coefficient estimates indicate more looks to the target object). Marginal effects in italics. No factors reached significance.

	SentenceStart to RCStart				RCStart to RCNounStart				RCVerbStart to RCNounStart			
	Objects		Agents		Objects		Agents		Objects		Agents	
Predictors	$\hat{\beta}$	p	$\hat{\beta}$	p	$\hat{\beta}$	p	$\hat{\beta}$	p	$\hat{\beta}$	p	$\hat{\beta}$	p
Intercept	0.49	0.11	-0.07	0.37	0.09	0.57	0.17	*0.06*	0.23	0.30	0.27	*0.07*
Cue (Cued=1 vs Uncued=-1)	0.27	0.33	-0.04	0.56	-0.17	0.30	-0.06	0.56	-0.22	0.35	-0.08	0.63
Phase (End=1 vs Beginning=-1)	-0.12	0.71	0.00	0.97	0.20	0.25	0.14	0.15	0.42	0.14	0.14	0.30
Cue:Phase	-0.27	0.45	0.04	0.57	0.01	0.93	-0.16	*0.10*	0.07	0.80	-0.25	*0.08*

til then. This meant that participants did not derive any time benefit from correctly tracking the talker's attachment preferences. However, we note that in the original experiment, as participants were asked only to look at the scene while listening to the sentence, there was also no explicit incentive to track the talker's attachment preferences.

An alternate possibility is that listeners do not exhibit talker-specific adaptation to attachment structure because they are not capable of it. Why might this be the case, particularly given the growing body of evidence for talker-specific *phonetic* adaptation? As talkers vary systematically in how they sound, both due to physiological factors (e.g. vocal tract size), and sociolinguistic factors (e.g. regional dialect), phonetic adaptation to talker-specific pronunciations can lead to a large benefit in later encounters with that same talker (Bradlow & Bent, 2008).

By contrast, syntactic adaptation, particularly syntactic adaptation to high and low attachment structure, may not carry the same utility: subject relative clauses alone occur fewer on than 2% of all noun phrases (Roland, Dick, & Elman, 2007). Non-local subject relative clauses with two animate (potential) heads—like those used in our experiment—are even rarer. Thus, talker-specific syntactic adaptation to such a low-frequency structure may not lead to the same benefits of improved language understanding as phonetic adaptation to much more frequent phonemes. It may be beneficial to address questions of talker-specific adaptation by examining more frequent syntactic structures.

In sum, the current experiment failed to replicate the effect that listeners are able to track talker-specific syntactic preferences for attachment (Kamide, 2012). This raises questions about the extent to which syntactic adaptation may involve the same adaptive mechanisms used in phonetic adaptation.

Acknowledgments

We thank Yuki Kamide for generously providing her visual stimuli and sentences, and for helpful discussion. We thank Anne Pier Salverda for technical and design advice, and members of HLP and KurTan lab at Rochester for feedback. This work was funded by NIH R01 HD075797 to TFJ. Views expressed here do not necessarily reflect those of the NIH.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *J. Acoust. Soc. Am.*, *119*(4).

Fine, A., Jaeger, T. F., Farmer, T., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, *8*(10).

Kamide, Y. (2012). Learning individual talkers structural preferences. *Cognition*, *124*(1), 66–71.

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

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *J. Mem. Lang.*, *56*(1), 1–15.

Kurumada, C., Brown, M., & Tanenhaus, M. (2012). Prosody and pragmatic inference: It looks like speech adaptation. In *Proc. 34th ann. conf. cognitive science society.*

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic english grammatical structures: A corpus analysis. *JML*, *57*(3), 348–379.

Salverda, A. P., & Tanenhaus, M. K. (in press). The visual world paradigm. In A. de Groot P. Hagoort (Ed.), *Research methods in psychology.* John Wiley & Sons.

Tooley, K. (2009). Is syntactic priming in sentence comprehension really just implicit learning. In *Proc. 22nd cuny.*

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Lang. Cognitive. Proc.*, *27*(7-8), 979–1001.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychol.*, *58*(2), 250–271.