Title

Memory maintenance of gradient speech representations is mediated by theirexpected utility

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Bushong, Wednesday
Jaeger, T. Florian

Publication Date
2019

Peer reviewed

Memory maintenance of gradient speech representations is mediated by their expected utility

Wednesday Bushong (wbushong@ur.rochester.edu) and T. Florian Jaeger (fjaeger@ur.rochester.edu)
Department of Brain and Cognitive Sciences, University of Rochester

Abstract

Language understanding requires listeners to quickly compress large amounts of perceptual information into abstract linguistic categories. Critical cues to those categories are distributed across the speech signal, with some cues appearing substantially later. Speech perception would thus be facilitated if *gradient sub-categorical* representations of the input are maintained in memory, allowing optimal cue integration. However, indiscriminate maintenance of the high-dimensional signal would tax memory systems. We hypothesize that speech perception balances these pressures by maintaining gradient representations that are expected to facilitate category recognition. Two perception experiments test this hypothesis. Between participants, an initial exposure phase manipulated the utility of information maintenance: in the *High-Informativity* group, following context always was informative; in the *Low-Informativity group*, following context always was uninformative. A subsequent test phase measured the extent to which participants maintained gradient representations. The Low-Informativity group showed less maintenance, compared to the High-Informativity group (Experiment 1). We then increased the task demands and made the targets of the manipulation less obvious to participants (Experiment 2). We found a qualitatively similar pattern. Together, these results suggest that listeners are capable of allocating memory to gradient representations of the speech input based on the expected utility of those representations.

Keywords: speech perception; cue integration; memory; expected utility

Introduction

Spoken (and signed) language is a temporally unfolding signal. In order to comprehend language, humans must quickly compress kilobits of information per second into abstract linguistic representations and meanings that contain more manageable amounts of information. At the same time, cues to linguistic categories often do not temporallly co-occur in neatly delimited segments of the speech signal, but rather are distributed across the signal. For example, one of the primary cues to stop voicing in English is the duration of the *preceding* vowel (Klatt, 1976). To make optimal categorization judgments, listeners must retain some sub-categorical information about the preceding vowel in memory in order to integrate it with later-arriving information (i.e., the stop itself). This kind of information distribution is typical across languages and can occur at several timescales: cues to sound categories can come not only from proximate acoustic properties, but also from, e.g., later lexical and semantic context that can occur anywhere. But maintaining rich representations of all incoming input would seemingly overload working memory. Thus, many theories of language processing claim that listeners simply do not maintain gradient representations of the input on any significant timescale, but instead immediately compress input into abstract representations (Just & Carpenter, 1980; Christiansen & Chater, 2016). According to these accounts, listeners throw away rich representations of the input as soon as a categorical perceptual judgment has been made.

However, a growing body of literature has suggested that listeners can and do maintain sub-categorical representations of prior input (McMurray, Tanenhaus, & Aslin, 2009), even at quite long perceptual timescales (Connine, Blasko, & Hall, 1991; Brown-Schmidt & Toscano, 2017; Gwilliams, Linzen, Poeppel, & Marantz, 2018). For example, Connine et al. (1991) exposed participants to sentences like "When the ?ent in the [fender/campground]...", where the ?-segment ranged between /d/ and /t/ (by manipulating one of the primary cues to voicing perception, the voice-onset time or VOT). The context following the ?-segment contained additional semantic context toward the identity of the original word. Participants had to categorize whether they had heard the word "tent" or "dent" in the sentence. Connine and colleagues found that participants' categorizations were influenced *both* by the VOT of the sound *and* by subsequent context, suggesting that listeners maintained a gradient representation of the initial sound for later use in cue integration and categorization. Subsequent studies have confirmed that listeners can maintain sub-categorical representations well beyond word boundaries (Szostak & Pitt, 2013; Bushong & Jaeger, 2017; Bicknell, Bushong, Tanenhaus, & Jaeger, under review).

How is this possible when language contains too much information to be held in memory indefinitely? We hypothesize that listeners use a memory strategy based on *expected utility*: the more important a piece of input is deemed to be, the more likely a detailed gradient representation should be maintained in memory; the less important the input, the more likely a categorical, less detailed representation will be maintained. In the case of the VOT and subsequent context example above, we can operationalize utility as the likelihood that *subsequent* context will be informative for categorization of the current input—if there is likely to be later information relevant to categorization, listeners should maintain a gradient representation of the speech input in order to be able to use it during cue integration (when the relevant subsequent context arrives).

In order to test this proposal, we conduct two experiments where we manipulate the probability that subsequent content in the sentence is relevant to the target word that participants have to categorize. Participants listen to sentences where a critical target word is acoustically manipulated to range between *tent* and *dent*. Like in the experiments by Connine and colleagues, these words are embedded in sentences. Unlike in earlier work, one group of participants hears sentences that always contain subsequent contextual information that is *informative* for categorization, whereas another group of participants hears sentences with *uninformative* subsequent context. Following exposure, we then test how much participants in the two groups maintain sub-categorical representations about the target word. Here, we operationalize whether participants are maintaining sub-categorical representations of the initial target as the extent to which each group integrates *both* acoustics of the target word *and* subsequent context into their categorization responses.

General Methods

Participants

We recruited 128 native English-speaking participants each for Experiments 1 and 2. Participants were recruited from Amazon Mechanical Turk and rewarded $3.00 for their participation. Participants could only participate in in either Experiment 1 or Experiment 2. The average age of our participants was comparable across experiments suggesting they had similar amounts of language experience (Experiment 1: 37.55 ± 12.04; Experiment 2: 34.14 ± 8.11).

Materials

We take the paradigm from Bushong and Jaeger (2017) as a starting point for our experiments. We constructed 12 sentence triplets like the following:

(1) After the ?ent Sue had found in the campgrounds collapsed, we went to a hotel. **(*tent*-biasing context)**

(2) After the ?ent Sue had found in the teapot was noticed, we threw it away. **(*dent*-biasing context)**

(3) After the ?ent was noticed, we continued on our way. **(neutral context)**

We manipulated two aspects of the sentence stimuli. First, we acoustically manipulated the "?" to range between /d/ and /t/ by changing the value of its voice-onset time (VOT), the primary cue distinguishing voiced from voiceless syllable-initial stop consonants in English. Based on norming and previous experiments, we chose to test VOT values of 10, 40, 50, 60, 70, and 85ms to cover a perceptual range from /d/ to /t/ with ambiguous points in between. Each VOT step occurred equally often. Second, we manipulated whether later context biased toward a /t/-interpretation (1) , /d/-interpretation (2), or neither (3). Informative words in the subsequent context—if present—occurred between 6-9 syllables after the target word, as in (1) and (2) above.

Procedure

Both experiments consisted of two phases (participants were unaware of this implicit structure): Exposure (72 trials) and Test (48 trials). Participants were randomly assigned to one of two groups: Low-Informativity exposure and High-Informativity exposure. During exposure, the Low-Informativity exposure group only heard sentences with neutral subsequent context (e.g., sentence (3) above), such that the only relevant information to sound categorization was VOT. The High-Informativity exposure group, by contrast, always heard sentences that contained informative later context (split evenly between /t/-biasing and /d/-biasing contexts), as in previous studies. In the test phase, both groups heard sentences that contained informative later context (split evenly between /t/-biasing and /d/-biasing contexts). This allowed us to assess context effects during the test phase, following Connine et al. (1991). Figure 1 illustrates the design of both experiments.

Both during exposure and test, participants' task was simply to categorize whether they heard one of two alternative words after they heard the full sentence. In Experiment 1, participants always made judgments about our critical target words of interest—i.e., they were asked whether they heard "tent" or "dent" on every trial. In Experiment 2, on half of all trials, participants instead had to categorize another word in the sentence (e.g., for sentence (3) above they were asked whether they heard "way" or "day"). We motivate this difference in design after presenting Experiment 1.

Predictions

We analyze responses from the test phase. Specifically, we analyze the influence of VOT and subsequent context on categorization responses to assess whether listeners maintained gradient representations of VOT. If participants maintain sub-categorical information about the /t/ and /d/ in the target word *?ent* until the end of the sentence, we should see effects of both VOT and context. Critically, if listeners can monitor the utility of subsequent context for the target word, and if expectations about this utility affect the degree to which listeners maintain sub-categorical representations, we should see that the main effect of context is smaller in the Low-Informativity exposure group, compared to the High-Informativity exposure group. Note that observing a continuous effect of VOT is not sufficient to establish that listeners maintain gradient representations of VOT since this could still reflect initial deterministic categorizations; it is the ability for listeners to integrate continuous VOT information with later-arriving context that is critical (for more discussion of this point, see Bicknell et al., under review).

Experiment 1

The goal of Experiment 1 was to test the simplest version of our proposal, whether listeners can adapt their expectations about the utility of subsequent context and use them to guide whether to maintain gradient representations of initial acoustic input.

Analysis

Following previous work (Bicknell et al., under review; Bushong & Jaeger, 2017), we excluded participants whose

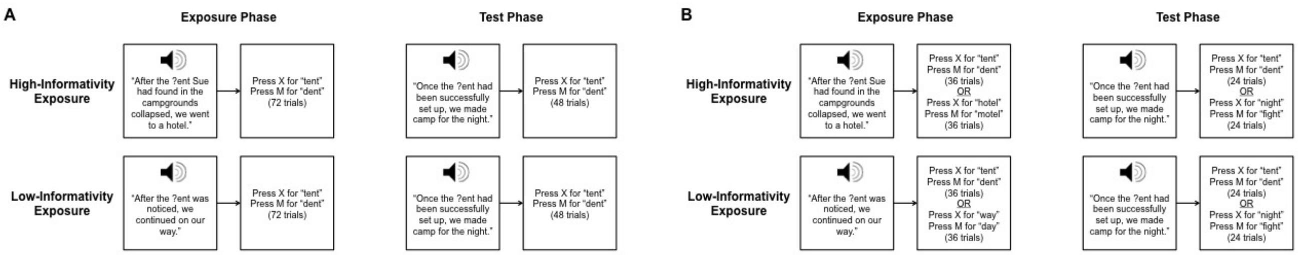

Figure 1: Design of the two experiments. **A**: Experiment 1. **B**: Experiment 2.

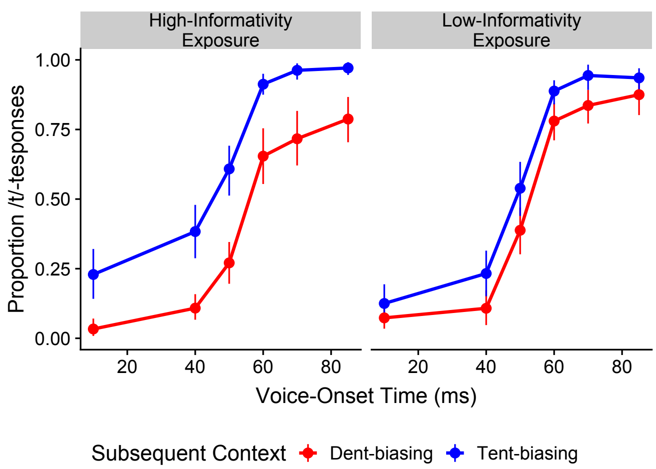

Figure 2: Experiment 1 test phase categorization results by VOT, subsequent context, and exposure group. Error bars are bootstrapped 95% confidence intervals over subject means.

categorization responses were not modulated by VOT, suggesting that they did not understand the task. This resulted in the exclusion of 11 participants for Experiment 1 (8.6%).

We fit mixed-effects logistic regression (Jaeger, 2008) predicting the proportion of /t/ responses in the test phase from VOT (z-scored to help with model convergence), squared VOT (z-scored), subsequent context (sum-coded: 1 = *tent*-biasing vs. -1 = *dent*-biasing), group (sum-coded: 1 = High-Informativity vs. -1 = Low-Informativity exposure), the two-way interaction of group with subsequent context, and the two-way interaction of subsequent context and squared VOT.[1] We included the interaction between squared VOT and context to test whether listeners' behavior was ideal observer-like (for a longer discussion on why this is important, see Bicknell et al., under review). The analysis also contained the full random effects structure that allowed model convergence. Analyses were conducted in the `lme4` package in `R` (Bates, Maechler, Bolker, Walker, et al., 2014).

Results

Figure 2 shows /t/-responses by group, VOT, and subsequent context over the test phase.

We found main effects of VOT ($\hat{\beta} = 2.87, p < 0.001$), squared VOT ($\hat{\beta} = 1.15, p = 0.01$) and subsequent context ($\hat{\beta} = 0.88, p < 0.001$) such that participants were more likely to respond /t/ when VOT increased and when subsequent context was /t/-biasing. The interaction between squared VOT and context was not significant ($\hat{\beta} = 0.06, p = 0.47$), replicating previous findings that suggest rational information integration.

There was no main effect of group on /t/-responses ($\hat{\beta} = 0.01, p = 0.89$). Crucially, there was a significant interaction between group and context ($\hat{\beta} = 0.41, p = 0.001$) such that the High-Informativity group showed a larger context effect than the Low-Informativity group. A simple effects analysis[2] revealed that both groups showed a significant context effect in the same direction (High-Informativity group: $\hat{\beta} = 1.25, p < 0.001$; Low-Informativity group: $\hat{\beta} = 0.47, p = 0.007$).

Discussion

We found that both exposure groups showed effects of VOT and subsequent context on their categorization responses, suggesting that they maintained gradient representations of speech input (VOT) in memory. As predicted, however, the effect of context was much smaller for the Low-Informativity group as compared to the High-Informativity group. These results suggest that the average informativity of later context influences whether listeners maintain gradient information about VOT in memory.

Experiment 1 shares with most previous work on the maintenance of sub-categorical representations that our paradigm involved a large degree of repetition (see Connine et al., 1991; Szostak & Pitt, 2013; Bicknell et al., under review; Bushong & Jaeger, 2017). This raises questions about the extent to which the results of Experiment 1 generalize to scenarios that more closely resemble the task demands of everyday language processing. In particular, participants in Experiment 1 were asked to make categorization judgments about the same critical target words of interest (*tent* and *dent*) throughout the entire experiment. This target word always occurred in

[1] Fixed effects R formula: `/t/-response ~ VOT + VOT`^{2} `Context + Group + Group:Context + VOT`^{2}`:Context`.

[2] R formula: `/t/-response ~ VOT + Group / Context` plus the same random effects as the main analysis.

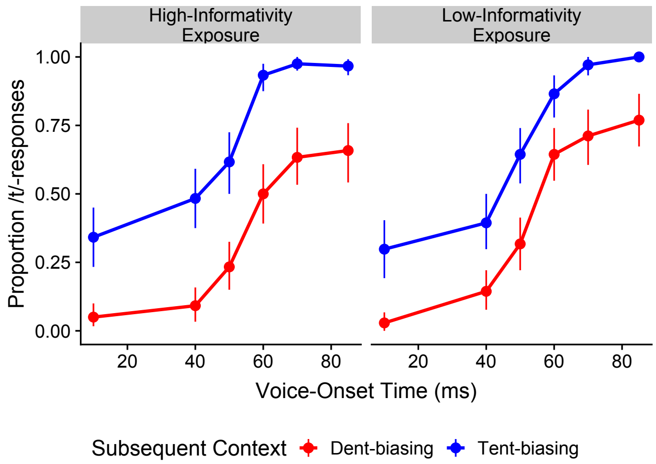

Figure 3: Experiment 2 Test Phase categorization results by VOT, subsequent context, and group. Error bars are bootstrapped 95% confidence intervals over subject means.

the same position within the 12 different sentence contexts. Additionally, subsequent information context (in the High-Informativity exposure group) always occurred about 6-9 syllables after the target. All of these factors likely directed participants' attention towards the target words.

One possibility is thus that the large context effects in the High-Informativity group is due to participants limiting their attention solely on the target word and the informative context. Similarly, the small context effect in the Low-Informativity group might be due to participants recognizing that they can do the task just as well while tuning out after hearing the target word, as they always get asked about the same target word. Experiment 2 presents a first step towards addressing this question.

Experiment 2

Experiment 2 is identical to Experiment 1, except that participants in both exposure groups of Experiment 2 only made judgments about the target words of interest (*tent* and *dent*) on half of the trials. On the other half of trials, participants were instead asked to categorize another word in the sentence; these alternate target words were never the informative subsequent context words for our critical words of interest—i.e., participants were never asked about the word "campgrounds" in sentence (1) above. This change from Experiment 1 had two purposes: (i) it directed participants' attention away from the target words, thus allowing us to test how participants behave when targets are not perfectly predictable, as in natural speech; and (ii) made it more likely that participants in both exposure groups remained attentive throughout the entire sentence. One of our concerns about Experiment 1 is that participants in the low-informativity group may have just 'tuned out' the rest of the sentence, and thus subsequent biasing context in the test phase, after hearing the target word. We reasoned that asking participants about words near the end of the sentence would generally increase attention toward those areas of the sentence, thus making it more likely that they heard and processed the later context. Participants were not told before the experiment what types of words they would be making judgments about or how often, so it is unlikely that they had a priori expectations about the distribution of target words (beyond general expectations about what kinds of words are usually tested in experiments, e.g. content words).

Analysis

Analyses were identical to Experiment 1. Our exclusion criteria resulted in the removal of 16 participants (12.5%) from analysis.

Results

Figure 3 shows /t/-responses by group, VOT, and subsequent context over the test phase.

We found main effects of VOT ($\hat{\beta} = 0.71, p = 0.01$), squared VOT ($\hat{\beta} = 1.68, p < 0.001$) and subsequent context ($\hat{\beta} = 1.26, p < 0.001$) such that participants were more likely to respond /t/ when VOT increased and when subsequent context was /t/-biasing. The interaction between squared VOT and context was significant ($\hat{\beta} = 0.22, p = 0.01$), in contrast to Experiment 1. In Experiment 2, we did not find an interaction between group and context ($\hat{\beta} = 0.18, p = 0.14$), although the numerical difference was in the same direction. The simple effects analysis revealed that both groups showed a significant context effect in the same direction (High-Informativity group: $\hat{\beta} = 1.39, p < 0.001$; Low-Informativity group: $\hat{\beta} = 1.01, p < 0.001$).

Discussion

The results of Experiment 2 are qualitatively similar to Experiment 1: participants overall showed evidence of integration of VOT and context into their responses, but the effect of context was numerically smaller in the Low-Informativity group than in the High-Informativity group. In contrast to Experiment 1, this difference between groups was not significant. This may suggest that shifting participants' attention away from our main manipulation made it harder to track how informative subsequent context was for our target words. However, it is hard to draw firm conclusions from a null result. To investigate whether the context effect difference is different between the two experiments, we directly compare them.

Interestingly, we also found a significant interaction between squared VOT and context, suggesting that participants' integration of VOT and context was non-optimal—unintuitively, participants seemed to use context more for *unambiguous* stimuli. This seems to contradict previous proposals that context is either integrated as a constant regardless of ambiguity (optimal integration, Bicknell et al., under review), or is used more for ambiguous than unambiguous stimuli (Connine et al., 1991). Since this aspect was not the focus of this experiment we will not discuss it further here, but further work should investigate why we observe suboptimal integration behavior in this more naturalistic setting

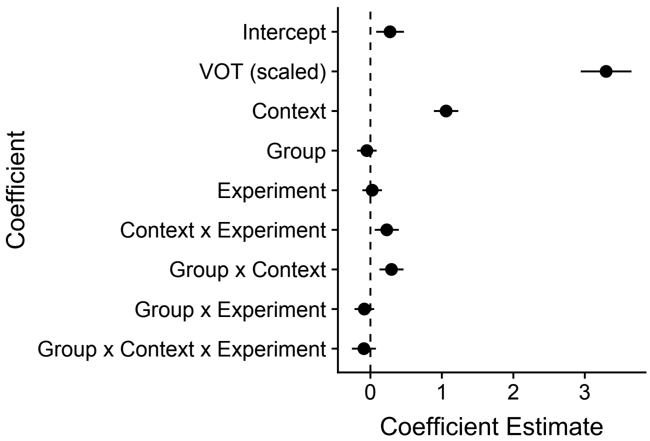

Figure 4: Model coefficients from combined analysis of Experiments 1 and 2. Error bars are 95% confidence intervals.

(for a first step in modeling sub-optimal integration strategies in this paradigm, see Bushong & Jaeger, 2019).

Comparison of Experiment 1 & 2

Since we observed a different pattern of results in Experiments 1 and 2, we conducted a post-hoc combined analysis in an attempt to assess overall evidence for the group by context interaction across experiments. In order to test the relationship between the group and context interaction and experiment, we fit a combined regression model to both of the datasets, allowing experiment (sum coded such that Experiment 1 = -1, Experiment 2 = 1) to interact with context and group. The other fixed effects remained the same as the above analyses for both experiments.

Results

Figure 4 shows the results of the combined analysis of Experiments 1 and 2.

In the combined analysis, we found main effects of VOT ($\hat{\beta} = 3.3, p < 0.001$) and subsequent context ($\hat{\beta} = 1.06, p < 0.001$). There was also a significant interaction between group and context ($\hat{\beta} = 0.29, p < 0.001$) such that the context effect was larger for the High-Informativity group than the Low-Informativity group. Critically, there was no three-way interaction between experiment, group, and context ($\hat{\beta} = -0.09, p = 0.29$), suggesting that the two-way interaction between group and context did *not* differ significantly across experiments.

We also found an unexpected two-way interaction between experiment and context ($\hat{\beta} = 0.23, p = 0.008$), such that the context effect was larger in Experiment 2 than Experiment 1. We return to this difference below. None of the other effects reached significance ($p > .2$).

General Discussion

When comprehending language, listeners must process thousands of bits of incoming information per second, compressing it into more manageable abstract representations. However, sub-categorical information about input can be useful to maintain in memory for later integration with relevant cues. Previous work has shown that listeners seem to be able to maintain such gradient representations in memory for up to several seconds. Here, we asked how this is possible when maintaining gradient representations of all incoming input would presumably overload short-term memory. We proposed that these effects may be driven by the *expected utility* of maintaining such information. In the case of these experiments, we tested this by manipulating the informativity of subsequent context: we reasoned that if subsequent context is likely to be informative for phonemic categorization, then the utility of maintaining gradient representations of VOT is higher, since it will be available for cue integration.

In Experiment 1, we found that both experimental groups maintained gradient representations of VOT over the timescale tested in this experiment (6-9 syllables). In line with our predictions, this effect was significantly smaller in the Low-Informativity group compared to the High-Informativity group. This provides support for our hypothesis that expected utility mediates maintenance of gradient representations of speech input in memory.

In order to make our experiments more naturalistic, we added an additional manipulation in Experiment 2. Participants were not always making judgments about our critical target words of interest. Instead, on half of all trials (during exposure and test), they made categorization judgments about non-critical words in the sentence. This change in the design takes a (small) step towards the task demands of natural language use: listeners don't necessarily *a priori* know which parts of the speech input they will need to comprehend and respond to. When we made this change, we observed the same numerical trend toward a smaller context effect in the Low-Informativity group, but this difference was not significant ($p = 0.14$). It is possible our manipulation in Experiment 2 successfully directed participants' attention away from the *tent*- and *dent*-biasing context, so that participants had a harder time estimating the informativity of subsequent context.[3] A follow-up combined analysis found no evidence that there was a difference in the interaction between the two experiments, though such an interaction might have been difficult to detect. We tentatively conclude that both experiments support the hypothesis that listeners maintain gradient representations according to their expected utility, but further experimentation with similar design is needed.

Of note, our follow-up analysis also found that participants in Experiment 2 exhibited an even *larger* context effect as compared to participants in Experiment 1. This might be seen as surprising: if anything Experiment 2 directed atten-

[3]As suggested by Figures 2 and 3, the difference in the context effects between experiments was driven by the Low-Informativity group: post-hoc analyses revealed that the context effect was larger in the Low-Informativity group in Experiment 2 compared to Experiment 1, but there were no differences in the High-Informativity group.

tion *away* from the critical word, compared to Experiment 1. In Experiment 2, participants had to make categorization judgments about words that occurred in different parts of the sentence rather than only the *tent* and *dent*. Critically, participants were never asked to make judgments about the *tent*- and *dent*-biasing context. The large context effect in Experiment 2 would thus seem to suggest that maintenance of gradient representations in memory is the default during speech perception, rather than the exception. While this interpretation stands in stark contrast with received wisdom (Just & Carpenter, 1980; Christiansen & Chater, 2016), it is line with a number of other recent findings that have found maintenance, for example, on the first trial of experiments (Bushong & Jaeger, 2017) or for lexically heterogeneous stimuli in naturalistic task-based language use (Burchill, Liu, & Jaeger, 2018; Brown-Schmidt & Toscano, 2017). Thus the present results may suggest that 'turning off' maintenance, rather than continued maintenance of gradient representations, requires sustained attention to specific, known targets.

Why would listeners show such a robust maintenance effect? One possible explanation consistent with our expected utility proposal is that natural language is *typically* informative: not only are low-level features like acoustic cues highly correlated even at long distances (providing helpful redundancy in light of perceptual inferences over noisy input, Hermansky, 2018), speakers talk about coherent topics that naturally provide semantic context that adds categorization-relevant information about the speech signal. Given these long-distance informational dependencies, maintaining gradient representations will typically be beneficial since it allows for optimal integration of these cues (Bicknell et al., under review). This would explain why we observe robust maintenance effects in paradigms that test sentences which follow these general natural constraints.

The present paradigm shares some caveats with previous work (Bicknell et al., under review; Connine et al., 1991; Szostak & Pitt, 2013): Experiments 1 and 2 involve a high degree of repetition; very much unlike in everyday language use, participants had to categorize the same word dozens of times. While the results of Experiment 2 show that listeners *do* maintain gradient representations about the speech input even when the target word is not *perfectly* predictable, Experiment 2 still allowed listeners to limit their attention—and maintenance of gradient representations: the critical target words (*tent* and *dent*) were the target of categorization on half of all trials. It is thus an open question whether equally strong maintenance of gradient representations is observed when it is less clear which aspects of the speech signal will turn out to be particularly relevant later (for preliminary evidence, see Burchill et al., 2018; Brown-Schmidt & Toscano, 2017).

Acknowledgments

This work was partially funded by NIHCD R01 HD075797 (to T.F.J.) and an NSF NRT #1449828 fellowship (to W.B.). The authors would like to thank Evan Hamaguchi, Nicole Vieyto, and Chelsea Marsh for assistance with stimulus sentence creation and recording.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using eigen and s4. *R package version*, *1*(7), 1–23.

Bicknell, K., Bushong, W., Tanenhaus, M. K., & Jaeger, T. F. (under review). Listeners can maintain and rationally update uncertainty about prior words.

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228.

Burchill, Z., Liu, L., & Jaeger, T. F. (2018). Maintaining information about speech input during accent adaptation. *PloS one*, *13*(8), e0199358.

Bushong, W., & Jaeger, T. F. (2017). Maintenance of perceptual information in speech perception. In *Proceedings of the thirty-ninth annual conference of the cognitive science society.*

Bushong, W., & Jaeger, T. F. (2019). Modeling long-distance cue integration in spoken word recognition. In *Proceedings of naacl 2019 cognitive modeling and computational linguistics workshop.*

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*(1), 234–250.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*(35), 7585–7599.

Hermansky, H. (2018). Coding and decoding of messages in human speech communication: Implications for machine recognition of speech. *Speech Communication*, 112–117.

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434–446.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329–354.

Klatt, D. H. (1976). Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category vot affects recovery from lexical garden-paths: Evidence against phoneme-level inhibition. *Journal of memory and language*, *60*(1), 65–91.

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533–1546.