Exploring the Label Feedback Effect: The Roles of Object Clarity and Relative Prevalence of Target Labels During Visual Search

by

Katherine P. Hebert

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2019 by the
Graduate Supervisory Committee:

Stephen D. Goldinger, Chair
Corianne Rogalsky
Samuel M. McClure
Viridiana Benitez

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

The *label-feedback hypothesis* (Lupyan, 2007, 2012) proposes that language modulates low- and high-level visual processing, such as priming visual object perception. Lupyan and Swingley (2012) found that repeating target names facilitates visual search, reducing response times and increasing accuracy. Hebert, Goldinger, and Walenchok (under review) used a modified design to replicate and extend this finding, and concluded that speaking modulates visual search via template integrity. The current series of experiments 1) replicated the work of Hebert et al. with audio stimuli played through headphones instead of self-directed speech, 2) examined the label feedback effect under conditions of varying object clarity, and 3) explored whether the relative prevalence of a target's audio label might modulate the label feedback effect (as in the *low prevalence effect;* Wolfe, Horowitz, & Kenner, 2005). Paradigms utilized both traditional spatial visual search and repeated serial visual presentation (RSVP). Results substantiated those found in previous studies—hearing target names improved performance, even (and sometimes especially) when conditions were difficult or noisy, and the relative prevalence of a target's audio label strongly impacted its perception. The mechanisms of the label feedback effect—namely, priming and target template integrity—are explored.

To HB, for your unwavering love and support.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Stephen Goldinger, for his continued expert guidance through the maze of academia, as well as my committee members for the generous donation of their time, insightful comments, and support. This project would also not have been possible without the tremendous programming assistance and patience of Dr. Stephen Walenchok.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF FIGURES

You're in the produce department of your local grocery store, and your eyes rapidly scan the walls of fruits and vegetables in search of mango. You mutter, "mango, mango, mango..." to yourself as you search, and ultimately find it next to the watermelon. Did repeating the word "mango" help you to find your delicious target? More specifically, did speaking the concept name facilitate visual perception for mango, perhaps by priming the visual concept or by assisting in the rejection of distractors? The *label-feedback hypothesis* (Lupyan, 2007; 2012; Lupyan and Swingley, 2012) proposes that hearing or speaking object names aids visual detection through a dynamic interaction between linguistic representations and feature detectors. Lupyan and Swingley (2012) found that repeating target names facilitates visual search, reducing response times and increasing accuracy. Hebert, Goldinger, and Walenchok (under review) used a modified design and collected oculomotor evidence to replicate and extend this finding. Participants searched for images of real objects (e.g., a mango) against a background of other objects, while simultaneously speaking during visual search. Four within-subjects, blocked conditions were tested. In different blocks, participants either (1) repeated target names during search (*target* condition), (2) repeated nonwords during search (*nonword* condition), (3) repeated names of real-world objects that were not present in the display (*distractor-absent* condition), or (4) repeated names of objects that *were* present in the display (*distractor-present* condition). Results showed that search was fastest while people spoke target names, followed in linear order by the *nonword, distractor-absent,* and *distractor-present* conditions. Gaze fixation patterns suggested that language does not affect attentional guidance, but instead affects both distractor rejection and target appreciation. Hebert et al. (under review) ultimately suggested that language affects template maintenance during search, allowing fluent differentiation of targets and distractors.

In the present work, I detail a series of five experiments designed to further explore the extent of the label feedback effect and the conditions under which label feedback impacts visual perception. **Experiment 1** replicated the work of Hebert et al. (under review), with audio stimuli instead of self-directed speech as the only change. This experiment is an important first step in expanding the growing body of evidence for the label feedback effect, and more crucially provided a more flexible method to be applied in future experiments.

Experiments 2A and 2B then investigated new visual circumstances under which the label feedback effect may prove perceptually beneficial. Previous research suggests that visual search becomes increasingly difficult when similarity between the mental representation of the target and the actual target image is low (Hout and Goldinger, 2015), and that the label feedback effect can potentially mitigate some of this cost (Lupyan and Swingley, 2012). For objects viewed below perceptual threshold (either by mask or rapid presentation), the label feedback effect has also been shown to "jumpstart vision" and boost those objects into visual awareness (Lupyan and Ward, 2013). In the real world, objects vary greatly in perceptual clarity, perhaps due to viewing distance, partial occlusion by another object, or poor vision. Might the label feedback effect also boost perception for items with poor visual clarity during search? Experiments 2A and 2B therefore investigated the label feedback effect under varying clarity conditions—within subjects, stimuli were either clear (*no blur*), slightly blurry (*minimal blur*), or completely blurry (*full blur*). I used the same paradigm as that of Hebert et al. (under review), but since we are now focused on the potential perceptual benefits of label feedback rather than exhaustive attentional manipulations, the *distractor-absent* and *distractor-present* conditions were eliminated.

I utilized both traditional spatial visual search (Experiment 2A) and passive rapid serial visual presentation (RSVP; Experiment 2B), to explore this question. In RSVP, stimuli are quickly presented in the center of the screen, one image at a time, and participants may only indicate their decision (target present or target absent) after the entire stream has been presented. This eliminates the ability to terminate search before all stimuli have been viewed, and simplifies participants' decision making, isolating search to perceptual decisions. It is also worth noting that many standard visual search phenomena replicate under RSVP procedure (Hout and Goldinger, 2010, 2012; Williams, 2010). One caveat of RSVP is that it does not give reaction time data, because observers must wait for every object to be displayed before rendering a present/absent decision, but this also ensures that an observer's eyes must land on each item in the search array. This is a large advantage to RSVP: any observed outcomes can be isolated to perceptual effects, because each item will have been viewed and analyzed to at least some extent. While traditional spatial search is arguably more externally valid and more closely mimics real-world search, RSVP provides the addition of signal detection measures, which are derived from accuracy but are much more nuanced than accuracy alone. *Signal detection theory* allows for the measurement of an observer's ability to differentiate between task-relevant information and irrelevant and/or random noise (Green & Swets, 1966; McNicol, 2005). In this context, signal detection measures provide a means to determine whether label feedback (i.e., hearing the name of a search target) influences an observer's ability to make this discrimination.

I anticipated that we would observe main effects for reaction times (RTs) and accuracy, both for the audio manipulation (*target* vs. *nonword*) and for the clarity manipulation (*no blur* vs. *minimal blur* vs. *full blur*). In other words, there would be a label feedback benefit when comparing the *target* condition to the *nonword* condition across all clarity conditions, and overall performance would be best on *no blur* trials, with a performance cost on *minimal blur* and *full blur* trials. More importantly, I predicted that there would be an audio vs. clarity interaction, where any benefits of the label feedback effect would be the most prominent under conditions wherein perceptual discrimination is difficult. When perceptual information is limited, noisy, or ambiguous, we increasingly rely on top-down information to "fill in the blanks" and make perceptual classifications (Scocchia, Valsecchi, and Triesch, 2014). I therefore expected that differences between performance in the *target* condition and the *nonword* condition would be greatest in the *full blur* condition—wherein the target audio information would be the most helpful in making perceptual discriminations—followed by the *minimal blur* condition and then the *no blur* condition. Such a finding would constitute further evidence that language-based activation of mental representations provides a top-down "boost" to visual perception, and that this "boost" is especially beneficial when perceptual discrimination conditions are noisy or difficult.

Finally, **Experiments 3A and 3B** explored how the relative prevalence of a search target label impacts visual perception. These experiments drew inspiration from the *low-prevalence effect*, a well-documented phenomenon wherein observers are exceedingly more likely to miss targets that occur rarely relative to the same targets that occur frequently (Wolfe, Horowitz, and Kenner, 2005). Jeremy Wolfe and colleagues (2005) found that when targets occurred with a prevalence of 50%, observers failed to notice them only 7% of the time, but when target prevalence was lowered to 10% and 1%, miss errors rose to 16% and a 30%, respectively. This means, for example, that airport bag screeners become expert water-bottle detectors, because prohibited water bottles occur often, but screeners have very poor performance in detecting actual weapons, which (thankfully) occur very rarely (Wolfe, Brunelli, Rubinstein, and Horowitz, 2013). The low prevalence effect has proven to be a very persistent and stubborn phenomenon with potentially dire consequences. While it applies to participants in laboratory settings, it also holds true for real-world settings and for trained expert observers like Transportation Security Administration baggage screeners and medical imaging professionals (Evans, Birdwell, & Wolfe, 2013; Evans, Tambouret, Evered, Wilbur, & Wolfe, 2011; Wolfe, et al., 2013), and the effect persists even when observers are forced to slow search down or are allowed to correct search errors (Kunar, Rich, & Wolfe, 2010). Troublingly, recent research also shows that observers still fail to detect low-prevalence targets approximately 12% - 34% of the time *even when they look directly at them* (Hout, Walenchok, Goldinger, & Wolfe, 2015).

Given that a target's prevalence and label feedback both impact search performance, it is conceivable that they would modulate each other. In the present study, participants searched for the same two targets throughout the entire experiment. Two between-subjects conditions manipulated the prevalence of a target label, a *prevalence* condition and a *control* condition. In the *prevalence* condition, subjects heard one of the two target names played through headphones on 20% of trials, and the second of the two target names was heard 80% of the time. In the *control* condition, subjects heard each target name 50% of the time. These target audio labels proportionally matched the search target present on-screen. As in Experiment 2, this experiment utilized both traditional spatial visual search (Experiment 3A) and RSVP (Experiment 3B).

I expected that there would be an overall effect of *audio label–target image match* across conditions, where performance would be better when the target present in the search array and the target name audio match, compared to trials in which there is a mismatch. I anticipated that this effect would be strongest, however, for audio labels that occur with higher prevalence. For participants in the *control* condition, this means we would not expect to see any differences in overall search performance between the two targets and two audio labels. For the *prevalence* condition, I expected that search performance would be better for whichever target had the more-frequently-heard audio label. Crucially, however, I anticipated that there would still be a label feedback effect for audio labels that occur with low prevalence. Observing a label feedback effect even when the audio label is rarely heard would not only demonstrate the power of the label feedback effect and speak to its robustness, but it would also indicate that perceptually it is—at least in part—a persistent low-level effect resistant to outside influence.

EXPERIMENT 1

Experiment 1 sought to replicate the work of Hebert et al. (under review), which conceptually replicated the label feedback effect by having participants search for images of real items amongst distractor objects while simultaneously speaking aloud across four different speaking conditions. Results showed that search was fastest while people spoke target names, followed in linear order by the *nonword, distractor-absent,* and *distractor-present* conditions, and Hebert et al. (under review) ultimately suggested that language affects target template maintenance during search. The present experiment used the same paradigm, with audio stimuli instead of self-directed speech as the only change. To foreshadow, the previously-observed behavioral trends successfully replicated, demonstrating the robustness of the label feedback effect and allowing for the use of audio stimuli in subsequent paradigms.

Method

Participants

To determine an appropriate sample size for Experiment 1, I conducted a repeated-measures ANOVA a priori power analysis for four within-subjects measures in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). I used the data from Hebert et al.'s (under review) *Experiment 1*, as the planned methods and apparatuses for the current experiment mimic that of Hebert et al.'s *Experiment 1* nearly completely. I converted the effect size of partial $\eta^2 = .13$ reported by Hebert et al. to Cohen's $f = 0.38$ using the formula from Cohen (1988). I used $\alpha = 0.05$ and the recommended power of .80 (Cohen, 1988), and ignored correlation among repeated measures to be conservative. According to this analysis, a minimum sample size of 20 would give the desired power for accuracy measures. For RTs, Hebert et al. (under review) reported a main effect of speaking condition, $F(3, 147) = 6.64$, $p < .001$, partial $\eta^2 = .12$. Using the same parameters as above, this converts to Cohen's $f = 0.37$. This analysis determined that 22 participants would allow sufficient power for RTs.

Given that previous effect sizes for both RTs and accuracy are moderately strong (Cohen, 1988), the required sample size for replication was relatively small. For the present experiment, to be conservative, 93 participants were recruited from the Arizona State University Psychology 101 subject pool. The participants were given course credit for their participation. All were native English speakers, and had normal or corrected-to-normal vision by self-report. Nine participants were excluded from data analysis based on performance—outliers were identified as anyone whose average RTs or error rates were ≥ 2.5 standard deviations above or below the group mean on any of the four visual search conditions.

Apparatus

Stimuli were presented using Dell computers and 19-inch LCD monitors, and participants responded via keyboard. Data was collected on up to 10 computers simultaneously, each with identical hardware and software, all in the same testing room under consistent lighting conditions. The experiment was administered using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Each participant used over-the-ear headphones to hear the audio stimuli.

Stimuli and design

As in Hebert et al. (under review), all target names, distractor names, and nonwords were 1-3 syllables (approximately 25% one-syllable, 50% two-syllables, and 25% three-syllables). One- and two-syllable nonwords were borrowed from Goldinger (1998); trisyllabic nonwords had prefixes or suffixes added onto bisyllables from the same list. Most object pictures came from the "Massive Memory" database (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010, cvcl.mit.edu/MM/stimuli.html), with a few taken from Google image searches. All images were sized to 100x100 pixels. Altogether, there were 192 unique target objects and approximately 2,000 distractors. I utilized similarity ratings from a multidimensional scaling database (Hout, Goldinger, & Brady, 2014) for object categories, and conflicting objects/categories were never paired. To create the audio stimuli, I wrote a Python program that fed each object name/nonword through Google Text-to-Speech (Python Software Foundation, Wilmington, DE, USA, https://www.python.org/). Using Google Text-to-Speech ensured that the voicing, tone, and speed of all spoken words were significantly more consistent than if a human being were to voice record each word manually (and saved considerable programming time). The python program read the name from a comma-separated values (.csv) list, fed it through Google Text-to-Speech, and then saved the spoken audio file as an .mp3. See Appendix A for Python code.

Again, following Experiment 1 from Hebert et al. (under review), there were four blocked, within-subject conditions (*target, nonword, distractor-absent, and distractor-present*). In every condition, participants heard words (or nonwords) at a steady pace during search, played through over-the-ear headphones. The conditions were blocked and presented in random order. Each block consisted of 48 trials, and each trial contained a unique target that appeared only once during the entire experiment. Search displays had one target and 24 distractors, and each object was placed in a random position on the screen.

Procedure

On each trial, participants were instructed to search for a target object among distractor objects, following the procedure shown in Figure 1. Both the target and the audio word (or nonword) were displayed verbally, and the audio word/nonword then began repeating. The "target" and "while listening to" labels remained on-screen until the participant pressed "ENTER" to begin the trial, which immediately initiated a screen instructing them to "get ready," which lasted four seconds. The audio continued to repeat during this time; this is to ensure that each item was heard at least 3-4 times, in the event that targets were found right away. After the "get ready" display, the search array appeared. Participants were instructed to press "SPACE" when they found the target, as quickly as possible. Audio stopped playing when the spacebar was pressed. RTs were measured from the onset of search displays to the spacebar press. After each response, the search array disappeared and numbers appeared on the screen in locations corresponding to each object for one second. The participants were then given a choice between two numbers from the previous screen, with one representing the target location. Participants chose the correct number by pressing the "F" or "J" keys, and "correct" or "incorrect" feedback was given. There were eight practice trials at the beginning of the experiment, two per condition. Experiment 1 lasted approximately one hour, with a break halfway through.

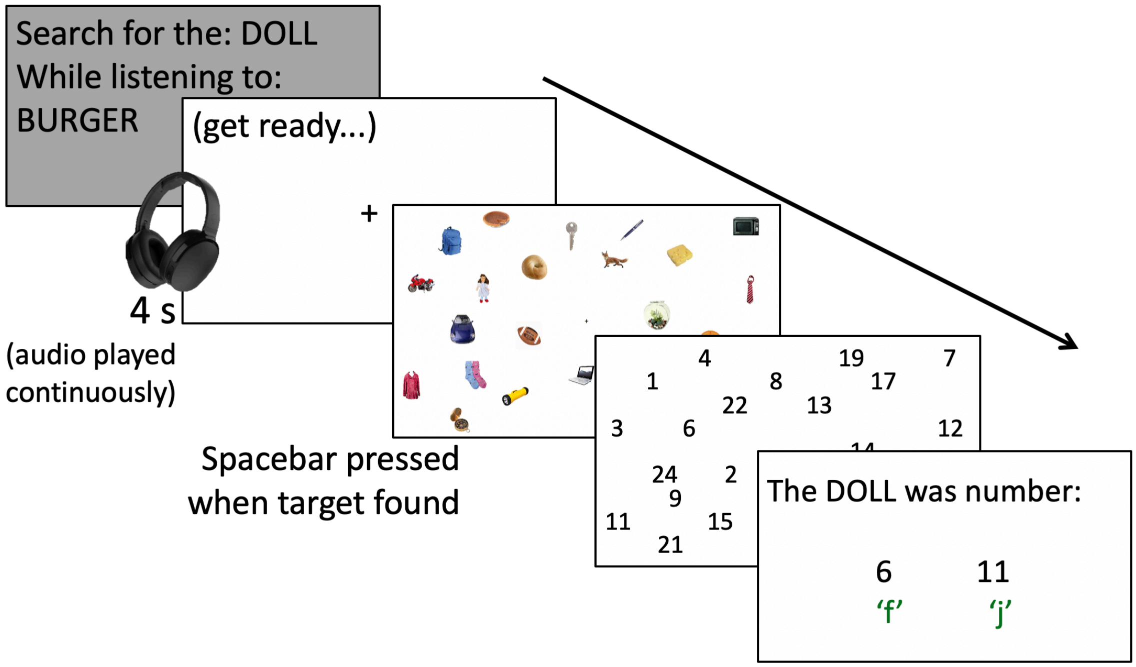

Figure 1. Procedure used in Experiments 1, 2A, and 3A.

Results

For accuracy and RTs, repeated-measures analyses of variance (ANOVAs) were performed with condition as a within-subject factor. Only correct responses were analyzed for RTs. Individual conditions were tested with planned paired comparisons (*t*-tests) using Bonferroni adjustments with a corrected alpha value of .05.

Participants in Experiment 1 were quite accurate across all four conditions (overall M = 97%). There was a main effect of audio label condition, $F(3, 83) = 3.008$, $p = .031$, partial $\eta^2 = .035$. Accuracy in the *target* condition (M = 97.6%) significantly exceeded the *nonword* condition (M = 96.6%; $t = 2.708$, $p = .008$), and the *distractor-present* condition (M = 96.6%; $t = 2.348$, $p = .021$), but did not differ significantly from the *distractor-absent* condition (M = 97.3%; $t = 0.72$, $p = .474$). There were no reliable differences (all $t < 1.6$) in accuracy among the latter three conditions. The RTs showed similar patterns: There was a main effect of condition, $F(3, 83) = 4.827$, $p = .003$, partial $\eta^2 = .055$. RTs in the *target* condition (M = 2,726 ms) were faster than the *nonword* condition (M = 2,885 ms; $t = 2.746$, $p = .007$), the *distractor-absent* condition (M = 2,961 ms; $t = 3.297$, $p = .001$), and the *distractor-present* condition (M = 2,904 ms; $t = 2.776$, $p = .007$). There were again no RT differences (all $t < 1.3$) among the latter three conditions (see Figure 2).

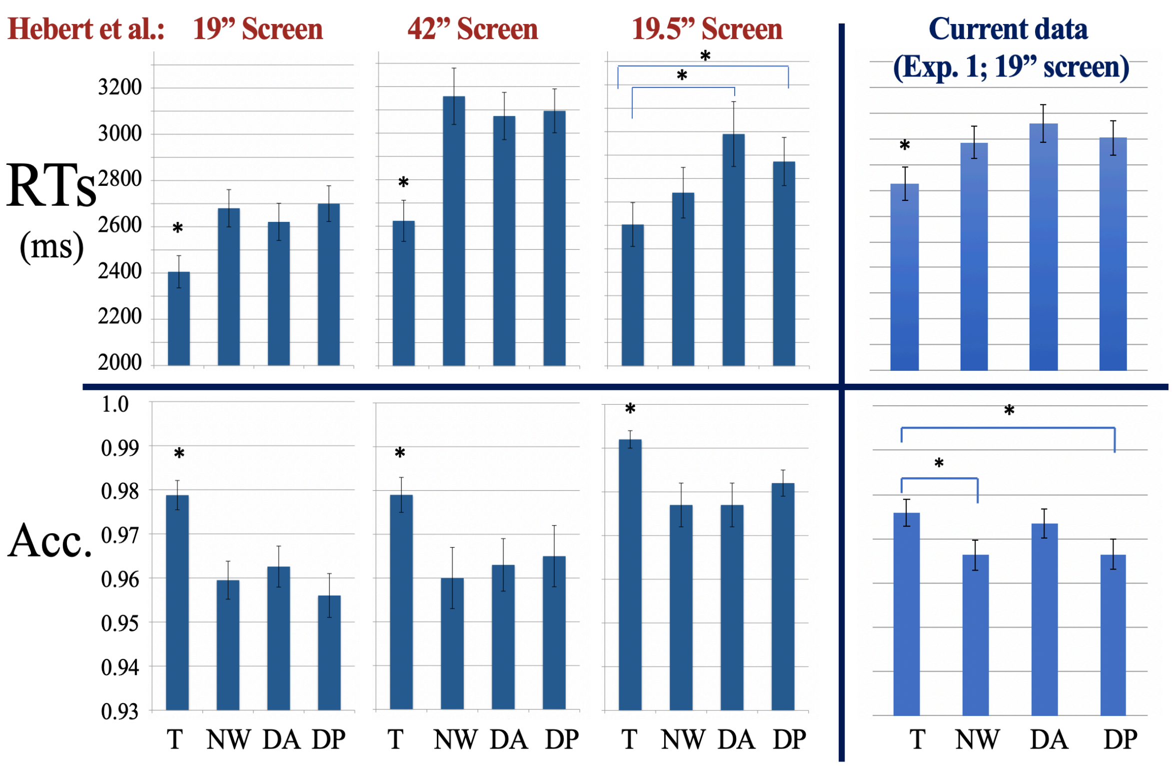

Figure 2. Search RTs (top) and accuracy (bottom) of Hebert et al. (under review; left) and the current experiment (right). In each panel, the conditions are shown: T = target; NW = nonword; DA = distractor-absent; DP = distractor-present (all trials). Error bars represent ±1 standard error of the mean (SEM).

Discussion

Experiment 1 sought to replicate the work of Hebert et al. (under review), using audio stimuli instead of self-directed speech. The same paradigm was used, with *target, nonword, distractor-absent,* and *distractor-present* audio played through headphones during visual search. Hebert et al. (under review) conducted three experiments in total, all of which conceptually replicated and extended Lupyan and Swingley (2012). The first two of these experiments measured just search RTs and accuracy, using 19-inch monitors (Hebert et al. *Exp. 1*) and 42-inch monitors (Hebert et al. *Exp. 2*). The third experiment utilized the same paradigm but incorporated eye tracking measures on 19.5-inch monitors (Hebert et al. *Exp. 3*). However, for the study at hand, replicating the overall behavioral trends is sufficient to determine whether a label feedback effect was indeed present with audio stimuli, as would be indicated by performance benefits in the *target* condition. The present experiment therefore examined only search RTs and accuracy, and utilized 19-inch monitors.

As in Hebert et al. (under review), the current results showed a clear label feedback effect, with a benefit of hearing the target name over hearing a nonword or a distractor word, both in accuracy and reaction time. The fact that there were no systematic behavioral differences between the *nonword, distractor-absent,* and *distractor-present* conditions is also consistent with previous findings. Eye movement analyses in previous research *did* demonstrate differences among those conditions, because eye-tracking allowed us to more closely examine the *distractor-present* condition and identify trials wherein the named distractor was (and was not) fixated during search. This was important because, if a named object was visible but never fixated, that trial is functionally equivalent to a *distractor-absent* trial. Without eye tracking, this differentiation is not possible (or, in this case, strictly necessary; see Figure 2 for comparison of behavioral results between the present study and previous research). For the sake of completeness, future studies could replicate the research with eye movement data to further solidify the relationship between label feedback and target template maintenance.

It is important to note that, while the results of Hebert et al. (under review) replicated consistently when changing from self-directed speech to audio stimuli, the results were not as robust. The overall partial η^2 for accuracy with self-directed speech was .13—which is a large effect size (Cohen, 1998)—whereas partial η^2 when using audio stimuli was only .035—a small effect size. For RTs, the partial η^2 with self-directed speech was .12 (large), and .055 (bordering medium) for audio stimuli. Additionally, in the current study, the accuracy of the *target* condition did not differ from that of the *distractor-absent* condition, whereas self-directed speech resulted in a significant difference between the two conditions. This and the difference in magnitude of effect sizes are not surprising, because the process of language production is much more effortful and susceptible to internal or external distraction than the process of language comprehension, which is generally considered to be more passive and automatic in most situations (Fedorenko, 2014). The fact that we observed a label feedback effect with audio stimuli in the current study is a testament to the effect's robustness, and adds to its growing body of evidence. Relevant to Experiments 2 and 3 in this paper, the present finding provides a better controlled and more flexible method of data collection that can be applied in future experiments.

EXPERIMENT 2

Experiments 2A and 2B investigated the label feedback effect under varying clarity conditions. Within subjects, stimuli were either clear (*no blur*), slightly blurry (*minimal blur*), or completely blurry (*full blur*). I used approximately the same paradigm as that of Hebert et al. (under review) and Experiment 1, but only the *target* and *nonword* conditions were included. Both traditional spatial visual search (Experiment 2A) and passive rapid serial visual presentation (RSVP; Experiment 2B) paradigms were used to explore this question. I anticipated that the label feedback effect would be largest in the full blur condition—wherein the target audio information would be the most helpful in making perceptual discriminations—followed by the minimal blur condition and then the no blur condition. This set of experiments provides evidence of new visual circumstances under which the label feedback effect seems to be perceptually beneficial.

Experiment 2A

Method

Participants. Seventy-two new participants were recruited from the Arizona State University Psychology 101 subject pool, and were given course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision by self-report. Six subjects were excluded from data analysis based on the same performance criteria as in Experiment 1.

Apparatus. As in Experiment 1, stimuli were presented using Dell computers and 19-inch LCD monitors, and participants indicated responses via keyboard press. Data was collected on up to 10 computers simultaneously, each with identical hardware and software, all in the same testing room under consistent lighting conditions. The experiment was administered using E-Prime 2.0 software (Schneider, Eschman, &

Zuccolotto, 2012). Each participant wore over-the-ear headphones to hear the audio stimuli.

Stimuli and design. Experiment 2A followed a 2 (audio label: *target* vs. *nonword*) x 3 (clarity: *no blur* vs. *minimal blur* vs. *full blur*) within-subjects design. *Target* and *nonword* conditions were blocked and counterbalanced by subject, with one block of 96 trials each. The clarity conditions occurred randomly throughout the entire experiment, with 32 trials of each clarity condition appearing in each block. In each trial, every stimulus in the search display—including the target—adhered to the same clarity condition (i.e., *no blur* stimuli will never appear on screen at the same time as *full blur* stimuli, etc.) In every condition, participants heard words (or nonwords) at a steady pace during search, played through over-the-ear headphones. In the *target* condition, the audio label always matched the name of the target object. Search displays contained one target and 24 distractors. Experiment 2A utilized traditional spatial search, and each object was placed in a random position on the screen.

For the *no blur* condition, Experiment 2A used the same object stimuli as that of the *target* and *nonword* conditions from Experiment 1. For the *minimal blur* and *full blur* conditions, the same stimuli were again used, but with a blur filter applied. To create the blurred stimuli, I applied a Python program to the file directory containing the unaltered images (Python Software Foundation, Wilmington, DE, USA, https://www.python.org/). For the *minimal blur* stimuli, the program applied a blur radius of one—wherein each pixel is set to the average value of the pixels in a square box extending one pixel in each direction—to each image, while saving that image under a new filename. The process was then repeated using a blur radius of three to achieve images with a *full blur*. See Appendix B for full Python code. I selected a blur radius of one for the *minimal blur* stimuli so that objects would still appear reasonably clear, but a small amount of perceptual information is lost. I used a blur radius of three for the *full blur* condition so that stimuli would appear the way they might if a visually-impaired person were to view an object without glasses—the edges of the image are still discernable, but the distinctions between many of the defining features of an object are often lost. See Figure 3 for examples of *no blur, minimal blur,* and *full blur* stimuli. For each of the three clarity conditions, there were 192 unique target objects to randomly select from, and approximately 2,000 distractors. Audio stimuli were the same as the stimuli in the *target* and *nonword* conditions from Experiment 1.

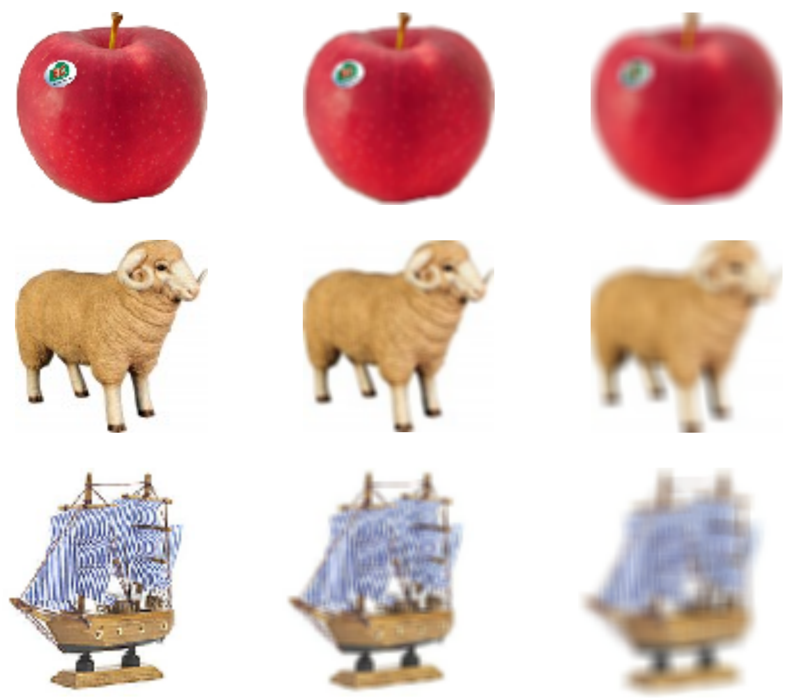

Figure 3. Example stimuli (top row: apple; middle row: sheep; bottom row: ship) from Experiment 2A and 2B's *no blur* condition (left), *minimal blur* condition (middle), and *full blur* condition (right).

Procedure. The procedure for Experiment 2A was nearly identical to that of Experiment 1, with minor changes. Instead of sometimes hearing the name of real-world distractor items (as in the *distractor-absent* and *distractor-present* conditions), participants heard either an accurate target label in the *target* condition or a nonword in the *nonword* condition. Because there were only two conditions in this experiment, there were fewer practice trials—participants completed six practice trials (instead of eight) at the beginning of the experiment, two per clarity condition. Experiment 2A lasted approximately one hour, with a break halfway through. All other procedural elements remained unchanged from Experiment 1.

Results

As in Experiment 1, repeated-measures analyses of variance (ANOVAs) were performed for accuracy and RTs, with audio label and clarity conditions as within-subject factors. Only correct responses were analyzed for RTs. Individual conditions were tested with planned paired comparisons (t-tests) using Bonferroni adjustments with a corrected alpha value of .05. Typical RT distributions are right skew, with a right tail of longer reaction times (McGill and Gibbon, 1963; Ratcliff and Murdock, 1976). The tail end of these distributions results in inflated RT means, with disproportionately high variance. To account for this issue, medians are used in these analyses rather than means. (Note that, since this approach was not used in Hebert et al. (under review), it was also not used in Experiment 1 analyses, which sought to replicate the original study closely. There were no qualitative differences between the two approaches, however. Going forward, medians will be used in all RT analyses.)

Participants in Experiment 2A were quite accurate across all audio and clarity conditions (overall M = 96%). There was a main effect of audio label condition, $F(1, 65)$

= 10.261, p = .002, partial η^2 = .136, indicating that accuracy in the *target* condition (M = 96.7%, SD = 3.42%) was significantly exceeded accuracy in the *nonword* condition (M = 95.1%, SD = 5.56%). There was also a main effect of clarity condition, $F(2, 130)$ = 15.392, p < .001, partial η^2 = .191. Accuracy in the *full blur* condition (M = 94.6%, SD = 5.59%) was significantly worse than in the *minimal blur* (M = 96.7%, SD = 3.83%; t = 4.205, p < .001) and *no blur* (M = 96.4%, SD = 4.19%; t = 5.108, p < .001) conditions. There was no significant difference between the *no blur* and the *blur* conditions, however, t = 0.91, p = .366.

There was a significant interaction between the audio label and blur conditions, $F(2, 130)$ = 3.13, p = .039, partial η^2 = .048. Mean differences in accuracy were largest between the *target* (M = 96.1%) and *nonword* (M = 93.2%) conditions with *full blur* stimuli, $t(65)$ = 3.28, p = .002. This difference was smaller but still present between the *target* (M = 97.4%) and *nonword* (M = 96.1%) conditions with *minimal blur* stimuli, $t(65)$ = 2.24, p = .029. During *no blur* stimuli trials, there was no significant difference between the *target* (M = 96.7%) and *nonword* (M = 96.0%) conditions, $t(65)$ = 1.093, p = .279 (see Figure 4A).

The RTs showed similar patterns: There was a main effect of audio label condition, $F(1, 65)$ = 6.603, p = .012, partial η^2 = .092, indicating that RTs in the *target* condition (M = 2,118 ms, SD = 728 ms) were significantly faster than in the *nonword* condition (M = 2,201 ms, SD = 740 ms; t = 2.57, p = .012). There was also a main effect of clarity condition, $F(2, 130)$ = 12.855, p < .001, partial η^2 = .165. RTs across *full blur* trials (M = 2,270 ms, SD = 824 ms) were significantly slower than in *minimal blur* (M = 2,129 ms, SD = 619 ms; t = 3.602, p = .001) and *no blur* (M = 2,080 ms, SD = 683 ms; t =

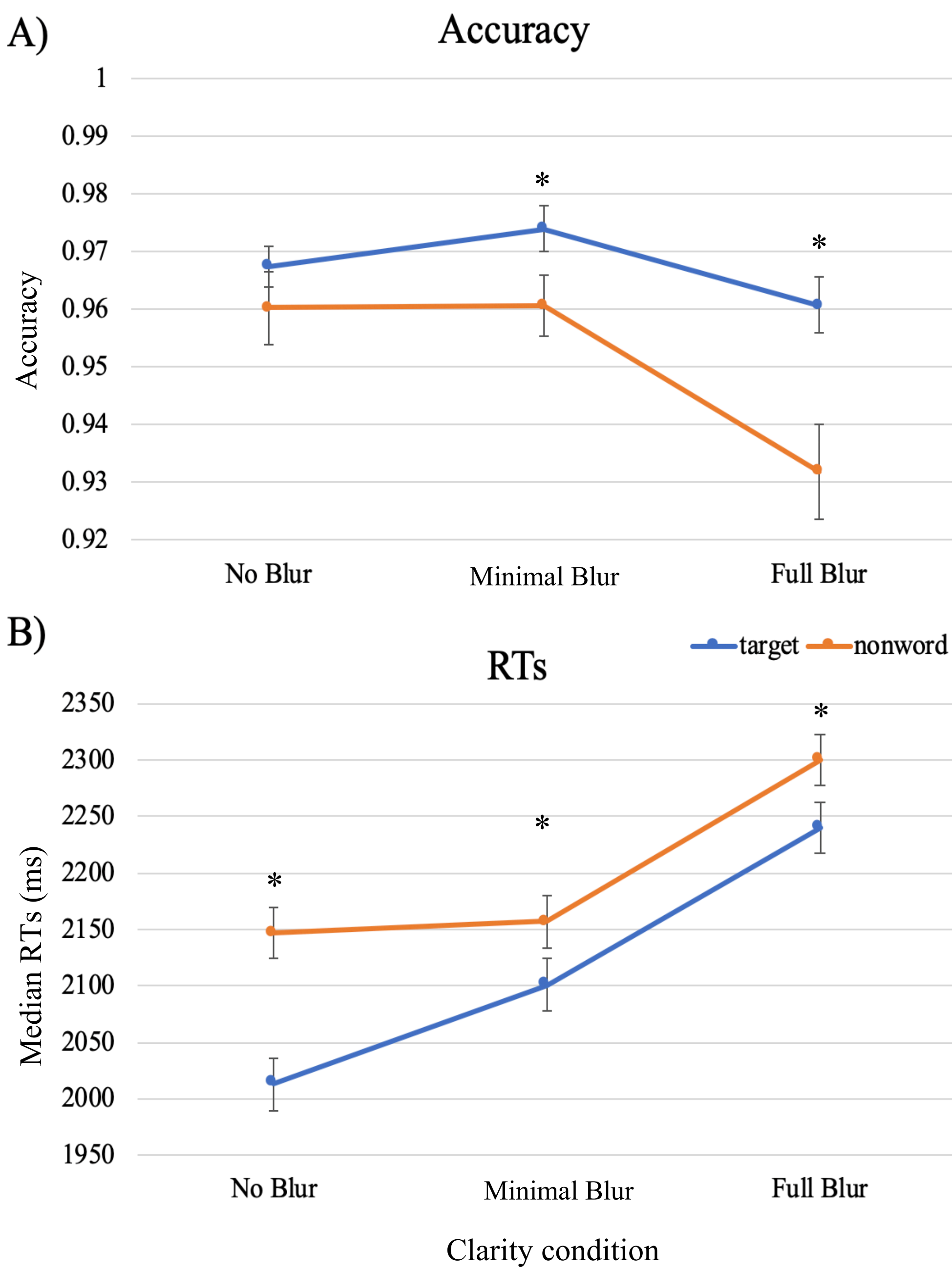

Figure 4. Experiment 2A accuracy (A) and reaction time (B) results. Error bars represent ±1 SEM.

4.891, $p < .001$) trials. There was no significant difference in RTs between *no blur* and the *minimal blur* trials, however, $t = 1.289$, $p = .59$.

There was no significant interaction between the audio label and blur conditions for RTs, $F(2, 130) = .405$, $p = .668$, partial $\eta^2 = .006$. Despite the lack of interaction, mean differences in RTs were significant between the *target* (M = 2,013 ms) and *nonword* (M = 2,147 ms) conditions for trials with *no blur* stimuli, $t(65) = 2.209$, p = .031, but not for the other two clarity conditions. Differences between audio label conditions was absent for trials with *minimal blur* stimuli (M_{target} = 2,101 ms; $M_{nonword}$ = 2,171 ms), $t(65) = 1.139$, $p = .259$, and for trials with *full blur* stimuli (M_{target} = 2,240 ms; $M_{nonword}$ = 2,361 ms), $t(65) = 1.618$, $p = .111$ (see Figure 4B).

Experiment 2B

Method

Participants. Fifty-seven new participants were recruited from the Arizona State University Psychology 101 subject pool, and were given course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision by self-report. Five subjects were excluded from data analysis based on the same performance criteria as in previous experiments.

Apparatus, stimuli, and design. Apparatus, stimuli, and design were identical to that of Experiment 2A, except for the fact that RSVP was used. Unlike spatial search (as in Experiment 2A), in RSVP, each object was displayed in the center of the screen one at a time. As in previous experiments, there were 24 distractor objects and one unique target on each trial, but the target was absent half of the time, in which case it was replaced by an additional distractor so that there were always 25 items in the search array. The target could not occur as either the first or last object in the display. The display of each object was followed by a backwards mask. Mask images are often used in RSVP—and in other similar visual tasks with brief stimulus exposure duration—to limit stimulus persistence (Felsten & Wasserman, 1980; Spencer & Shuntich, 1970). In the current experiment, the mask consisted of a square with wavy greens, browns, and pinks, and was sized to the same dimensions (100 x 100 pixels) as that of the object stimuli (see Figure 5 for a depiction of the mask image). The same mask was used on every trial throughout the experiment.

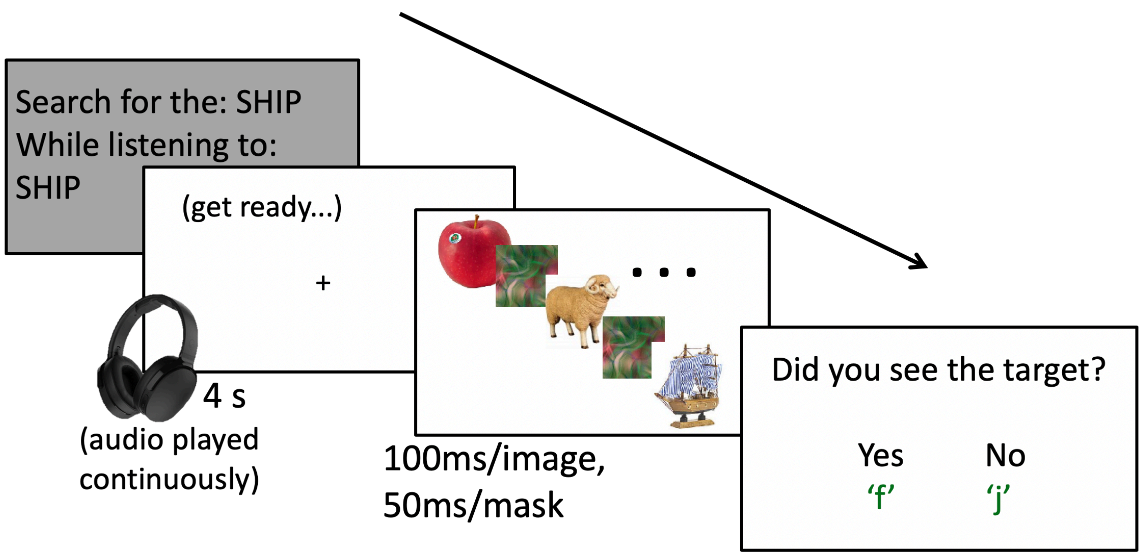

Figure 5. Procedure used in Experiments 2B and 3B.

Procedure. The RSVP version of Experiment 2 is procedurally very similar to that of the traditional spatial search version (Experiment 2A), again with relatively minor changes. On each trial, participants were instructed to search for a target object among distractor objects. Both the target and the audio word (or nonword) were displayed verbally on the monitor, and the audio word/nonword began repeating. The participant pressed “ENTER” to begin the trial, and there was again a four second “get ready” delay, during which the audio continued to repeat. At the onset of the search array, each of the 25 objects displayed rapidly in the center of the screen, one object at a time. Each item was displayed for 100ms, with a 50ms mask after each image. The nature of RSVP requires that the target be absent on 50% of trials, since it is impossible for a participant to indicate exactly where in the rapid search array an item occurred. Instead, after the final image disappeared, the audio stopped playing, and participants were then asked to indicate—by pressing the “F” or “J” keys—whether or not the target was present in the array (see Figure 5 for depiction of RSVP procedure). “Correct” or “incorrect” feedback was given after each trials. There were six practice trials at the beginning of the experiment, two per clarity condition. Experiment 2B lasted approximately one hour, with a break halfway through.

Results

The data were analyzed in an identical fashion to that of Experiment 2A, but only accuracy and subsequent signal detection measures were examined.[1] Individual conditions were tested with planned paired comparisons (*t*-tests) using Bonferroni adjustments with a corrected alpha value of .05.

Overall accuracy across all audio label and clarity conditions was 84.8%. There were main effects of audio label, $F(1, 51) = 4.406$, $p = .041$, partial $\eta^2 = .08$, and clarity, $F(2, 102) = 40.77$, $p < .001$, partial $\eta^2 = .44$, but there was no interaction, $F(2, 102) = .598$, $p = .55$, partial $\eta^2 = .01$. The main effect of audio label revealed that, across all

[1] Reaction time data does not exist in the typical sense for RSVP paradigms, since the participant must view all 25 objects before indicating whether she saw the target.

clarity conditions, accuracy in the *target* condition (M = 85.5%, SD = 4.75%) was significantly higher than in the *nonword* condition (M = 84.1%, SD = 4.71%, $t(51)$ = 2.099, p = .041. Performance was overall worse in *full blur* trials (M = 80.5%, SD = 5.48%) than in *minimal blur* trials (M = 86.7%, SD = 5.16) $t(51)$ = 8.02, p < .001, and in *no blur* trials (M = 87.1%, SD = 5.20%), $t(51)$ = 7.82, p < .001. There was no significant overall difference between *no blur* trials and *minimal blur* trials, $t(51)$ = .432, p = .667.

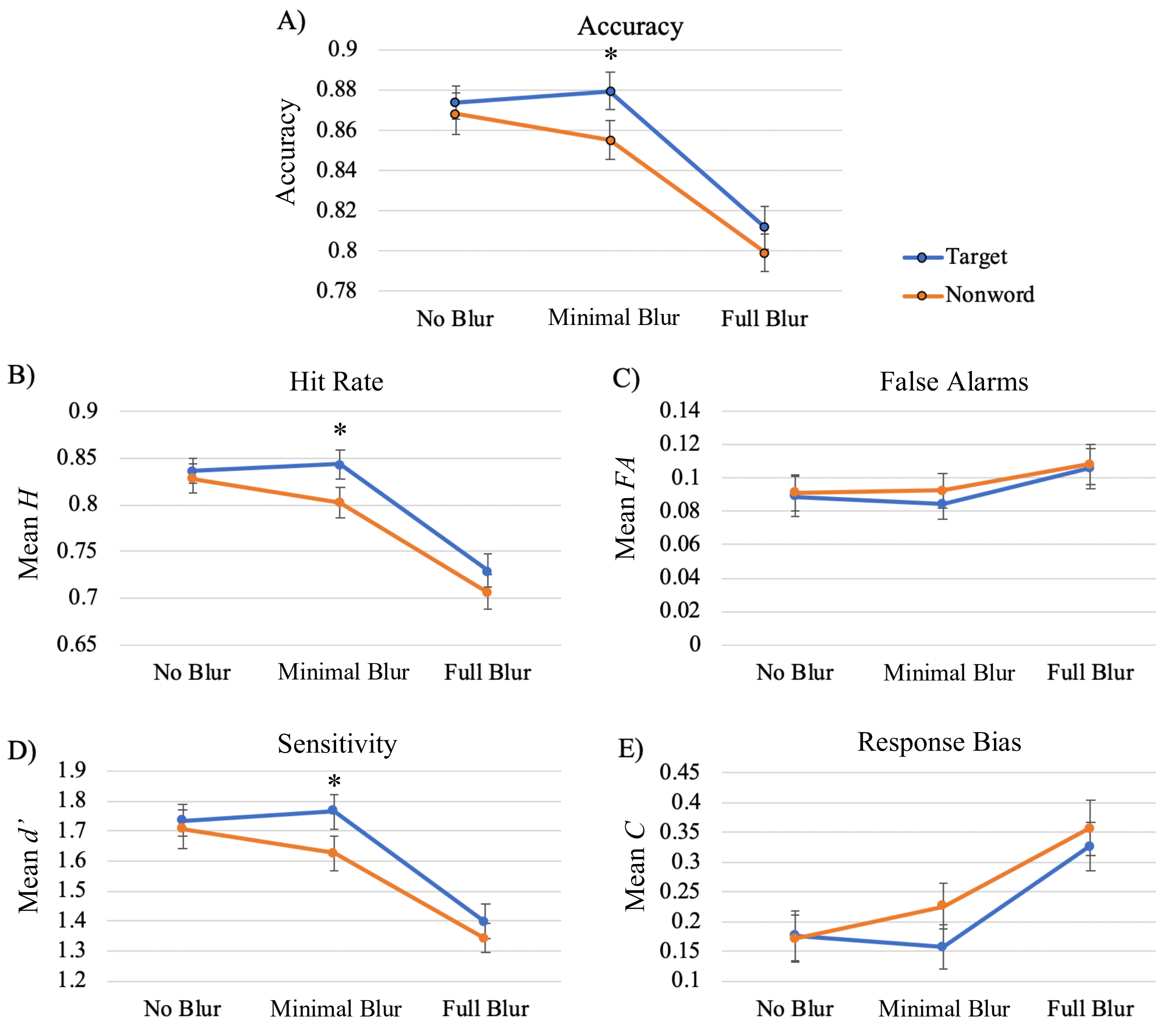

Figure 6. Experiment 2B results, with clarity condition plotted against audio label. Accuracy (A), hit rate (B), false alarm rate (C), d' (D), and C (E). Error bars represent ±1 SEM.

For trials with *minimal blur* stimuli, accuracy was significantly higher for the *target* condition (M = 88%) than in the *nonword* condition (M = 85.5%), $t(51) = 2.07$, $p = .044$. Differences between the *target* and *nonword* conditions for *no blur* and *full blur* trials were not significant (t's < .97; see Figure 6).

Accuracy was then further broken down into signal detection measures. *d'* (d prime), also known as sensitivity, reflects the standardized difference between the mean of the signal-present distribution and the mean of the signal-absent distribution, and it is calculated from a subject's hit proportion (H) and false alarm rate (FA; Green & Swets, 1966). When a subject made no false alarms (i.e., FA was 0.0), in order to calculate the z score (see Brophy, 1986), a standard correction (see Macmillan & Kaplan, 1985; Stanislaw & Todorov, 1999) was applied:

$$FA = \frac{1}{2N} \tag{1}$$

with N being the maximum number of false alarms. A similar standard correction was applied when a subject's H was 1.0:

$$H = 1 - \frac{1}{2N} \tag{2}$$

where N is the maximum number of hits. These corrections are akin to committing half of a false alarm and half of a miss. After calculating z scores for each participant's hit rate and false alarm rate, *d'* was then calculated using the following formula:

$$d' = \frac{1}{\sqrt{2}}[z_H - z_{FA}] \tag{3}$$

This formula takes into account the nature of a two-alternative forced-choice task, and differs from the *d'* formula used in old/new recognition memory paradigms (see Macmillan & Creelman, 2005).

A subject's criterion (*C*) refers to a subject's response bias, and reflects the extent to which one response (in this case, "target present" vs. "target absent") is more probable than the other (Green & Swets, 1966). *C* was calculated using the following formula:

$$C = -\frac{1}{2}\left[z_H + z_{FA}\right] \tag{4}$$

From there, ANOVAs and planned pairwise comparisons were used to analyze signal detection measures, with Bonferroni corrections applied.

A label feedback effect was present in hit rates and in *d'*. Collapsed across all clarity conditions, hit rates in the *target* audio label condition (M = .803) exceeded hit rates in the *nonword* condition (M = .779), $t(51) = 2.10$, $p = .040$. d'_{Target} (M = 1.63) was marginally larger than $d'_{nonword}$ (M = 1.55), $t(51) = 1.93$, $p = .059$. There were no differences in either *FA* or *C* between the *target* and *nonword* conditions, *t*'s < 1.08. There was a main effect of object clarity for both *d'*, $F(2, 51) = 16.85$, $p < .001$, partial $\eta^2 = .25$, and for *C*, $F(2,51) = 6.89$, $p = .002$, partial $\eta^2 = .12$. Across all measures, performance was worst in *full blur* trials, with little difference between *no blur* and *minimal blur* trials (see Figure 6). There was a significant clarity *x* audio label interaction for both *d'*, $F(2, 51) = 15.57$, $p < .001$, partial $\eta^2 = .23$, and for *C*, $F(2, 51) = 7.60$, $p = .001$, partial $\eta^2 = .13$; participants became more conservative (as indicated by a larger *C*) when visual conditions were difficult, but were less conservative overall in the *target* condition. The label feedback effect was once again largest for the *minimal blur* trials: Hit rates on these trials were significantly higher for the *target* condition (M = .844) than in the *nonword* condition (M = .803), $t(51) = 2.01$, $p = .050$, and d'_{target} (M = 1.77) was marginally larger than $d'_{nonword}$ (M = 1.62), $t(51) = 1.909$, $p = .062$. *FA* and *C* for *minimal blur* trials did not significantly differ from each other, however, *t*'s < 1.4. There was no significant evidence of increased false alarms to the named target regardless of object clarity. Differences in *H*, *FA*, *d'*, and *C* between the *target* and *nonword* conditions for *no blur* and *full blur* trials were all insignificant (t's < 1.01; see Figure 6).

Discussion

Experiments 2A and 2B investigated the label feedback effect under varying clarity conditions—within subjects, stimuli were either clear (*no blur*), slightly blurry (*minimal blur*), or completely blurry (*full blur*). Both traditional spatial visual search (Experiment 2A) and passive rapid serial visual presentation (RSVP; Experiment 2B) were used. I predicted that there would be a label feedback benefit when comparing the *target* condition to the *nonword* condition across all clarity conditions, and that overall performance would be best on *no blur* trials, with a performance cost on *minimal blur* and *full blur* trials. I also predicted that there would be an audio label *x* clarity interaction, with the strongest label feedback effect occurring under conditions wherein perceptual discrimination is difficult.

Results showed a clear label feedback effect evident throughout both experiments, manifesting either as a main effect or an interaction—consistent with previous research, hearing the name of the target object generally resulted in better search performance compared to hearing a nonword. Additionally, overall performance was indeed poorest in *full blur* trials, which demonstrates that the manipulation of stimuli clarity was successful. In spatial search, there was an interaction between audio label and clarity—differences in accuracy between *target* and *nonword* conditions were largest in *full blur* trials. This would seem to suggest that label feedback seems to be especially helpful in target detection when visual conditions are difficult (as suggested by Lupyan and Ward, 2013). However, this interaction is not statistically present in RTs, and in fact the largest

RT label feedback effect occurs in *no blur* trials, rather than in trials with blurry stimuli. While hearing the target name during search can certainly facilitate performance, perhaps a key mechanism of label feedback is the ability to *protect* against distraction and other performance-inhibiting factors—a minor but important distinction. Hebert et al. (under review) suggested that language affects distractor rejection, target appreciation, and target template maintenance during search, allowing fluent differentiation of targets and distractors. The current findings would seem to fit this narrative: hearing the name of the target object can indeed facilitate visual search, but hearing something other than the target name when visual conditions are difficult is especially detrimental to performance. Hearing the target name can protect the integrity of a target's template in working memory, but this template degrades when hearing a nonword, *especially* when visual input is dissimilar from this mental template (i.e., when object clarity is poor).

RSVP results (Experiment 2B) showed patterns similar to that of spatial search. Hit rates and false alarms followed similar trends to those observed in spatial search, with a main effect of object clarity. Interestingly, there was no significant evidence of increased false alarms to the named target, regardless of object clarity. This finding aligns with the theory that hearing a target label offers a level of protection against a degrading target template, as opposed to mainly facilitating search; if this wasn't the case, or if priming alone was the main mechanism behind the effect, then we would expect false alarms to increase in the *target* condition compared to the *nonword* condition. As predicted, sensitivity (*d'*) was larger in *no blur* and *minimal blur* trials than in *full blur* trials, especially when hearing the target name. As visual conditions became more difficult, *d'* diminished, but hearing the target name helped to preserve sensitivity over hearing a nonword. This was coupled with a criterion shift—participants became more conservative (as indicated by a larger *C*) when visual conditions were difficult, but were slightly less conservative when hearing the target name. So even though the label feedback effect did not lead to increased false alarms in the *target* condition, when compared to hearing nonwords, it did result in a slight shift towards a "target present" response criterion.

Experiment 2 also revealed an interesting trend wherein the mean accuracy (and subsequent signal detection measures) for *minimal blur* trials was nearly identical to that of *no blur* trials, but *minimal blur* trials saw a marked increase in the label feedback effect (see Figure 6). This was especially true under RSVP, where target detection must happen very rapidly, with no opportunity for refixations or for attention to be guided to the target. It could be argued that target template maintenance is even more crucial under these conditions, as there is no opportunity to "wander" and then "refocus" during a trial—if the target is missed during RSVP, there is no second chance to find it. This is evidenced by the fact that this phenomenon is present in measures of RSVP signal detection and is not present in spatial search RTs—target appreciation is the only factor in a two-alternative forced-choice task (as in RSVP), whereas RTs in spatial search are composed of both target appreciation (decision time) and attentional guidance (time to target fixation). The greater importance of uninterrupted target template maintenance in RSVP could potentially explain the reliable increase in the label feedback effect on *minimal blur* trials observed in Experiment 2B. *No blur* stimuli match the target template closely, and so this template is relatively robust, so that it does not require much protection and is not especially susceptible to degradation from hearing nonwords. For *full blur* trials, similarity between the target template and visual input is minimal, as visual conditions are difficult, and therefore the target template is simply not particularly helpful in assisting with target identification. Previous research has indeed demonstrated that as the discrepancy between a target template and visual input increases, the label feedback effect diminishes (Lupyan & Swingley, 2012), further evidencing this account. This brings us to the *minimal blur* stimuli, which match the target template reasonably well, but not entirely. It could be the case that these stimuli are in a "sweet spot" where they are similar enough to their target templates that hearing the target name still facilitates search, but visual conditions are *just difficult enough* that without hearing the target name, identifying the target object is much more difficult. This is evidenced by the fact that, between *no blur* and *minimal blur* trials, performance increases slightly during the *target* condition, and decreases during the *nonword* condition.

There remains the possibility that the increased label feedback effect in *minimal blur* trials could just be a fluke, of course—the fact that it repeats across all signal detection measures is not surprising, because they all are derived from overall accuracy. It is also possible that the levels of blur in the *minimal blur* and *full blur* stimuli would have been better-suited at a different blur radius, or that some other form of visual distortion would yield different results. Replication is needed, and further research could potentially manipulate visual noise and object clarity in order to model at what point the label feedback effect peaks and diminishes. Additionally, utilizing eye movement measures—specifically, parsing out attentional guidance and decision time—could shed light on this observed phenomenon.

EXPERIMENT 3

Previous research shows that the relative prevalence of a stimulus changes the way that stimulus is perceived (e.g., Hon & Tan, 2013; Laberge & Tweedy, 1964; Miller & Pachella, 1973; Wolfe et al., 2005). Experiment 3 explored whether the prevalence of an audio label might modulate the label feedback effect, or vice versa. Participants repeatedly searched for the same two targets, in both traditional spatial search (Experiment 3A) and RSVP (Experiment 3B), while consistently hearing the names of the targets. The prevalence of each target's audio label varied by condition. I anticipated that performance would be best on trials wherein the audio label matched the target present on the screen, but that the relative prevalence of the audio labels would modulate this performance difference.

Experiment 3A

Method

Participants. 181 new participants were recruited from the Arizona State University Psychology 101 subject pool, and were given course credit for their participation. For both Experiments 3A and 3B, the *prevalence* condition was oversampled because of the need to counterbalance the two target names played through headphones on 80% vs. 20% of trials, whereas in the 50/50 *control* condition no such counterbalancing was needed. Of the 181 participants, 52 were randomly assigned to *Group Muffin* (participants were not told this), and heard "muffin" on 80% of trials and "pinecone" on 20% of trials. *Group Pinecone* heard "pinecone" on 80% of trials and "muffin" on 20% of trials, and consisted of 70 participants. *Group Muffin* and *Group Pinecone* combined to give the *prevalence* condition 122 total participants. Lastly, the *control* condition consisted of the *balanced group*, which heard "muffin" and "pinecone" each on 50% of trials, and had 59 participants. All 181 participants were native English speakers and had normal or corrected-to-normal vision by self-report. Exclusion criteria was the same as in Experiments 1 and 2. After these criteria, six participants were excluded (four from *Group Muffin*, one from *Group Pinecone*, and one from the *balanced group*) resulting in a total n of 175: 117 in the oversampled *prevalence* condition and 58 in the *control* condition.

Apparatus. As in previous experiments, stimuli were presented using Dell computers and 19-inch LCD monitors, and participants responded via keyboard. Data was collected on up to 10 computers simultaneously, each with identical hardware and software, all in the same testing room under consistent lighting conditions. The experiment was administered using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Each participant wore over-the-ear headphones to hear the audio stimuli.

Figure 7. Target stimuli from Experiment 3. The muffin (left) and pinecone (right) each appeared on 50% of all trials.

Stimuli and design. The same two targets were used throughout all conditions and across all subjects. These targets were a muffin and a pinecone, and—like all distractor stimuli—were sized to 100 x 100 pixels (see Figure 7). Each target was present in the search array 50% of the time. There were two between-subjects conditions, a *prevalence* condition and a *control* condition. In the *prevalence* condition, one of the two target names was played through headphones on 80% of trials (the *high prevalence target audio label*), and the second of the two target names was heard 20% of the time (the *low prevalence target audio label*; see Figure 8). As noted previously, the target heard on 80% of trials vs. 20% of trials was counterbalanced by participant. This means that in the *prevalence* condition, the *high prevalence audio label* matched the target image 40% of the time (e.g., muffin present on 50% of trials *x* "muffin" heard on 80% of trials = 40% match), and the *low prevalence audio label* matched the target image only 10% of the

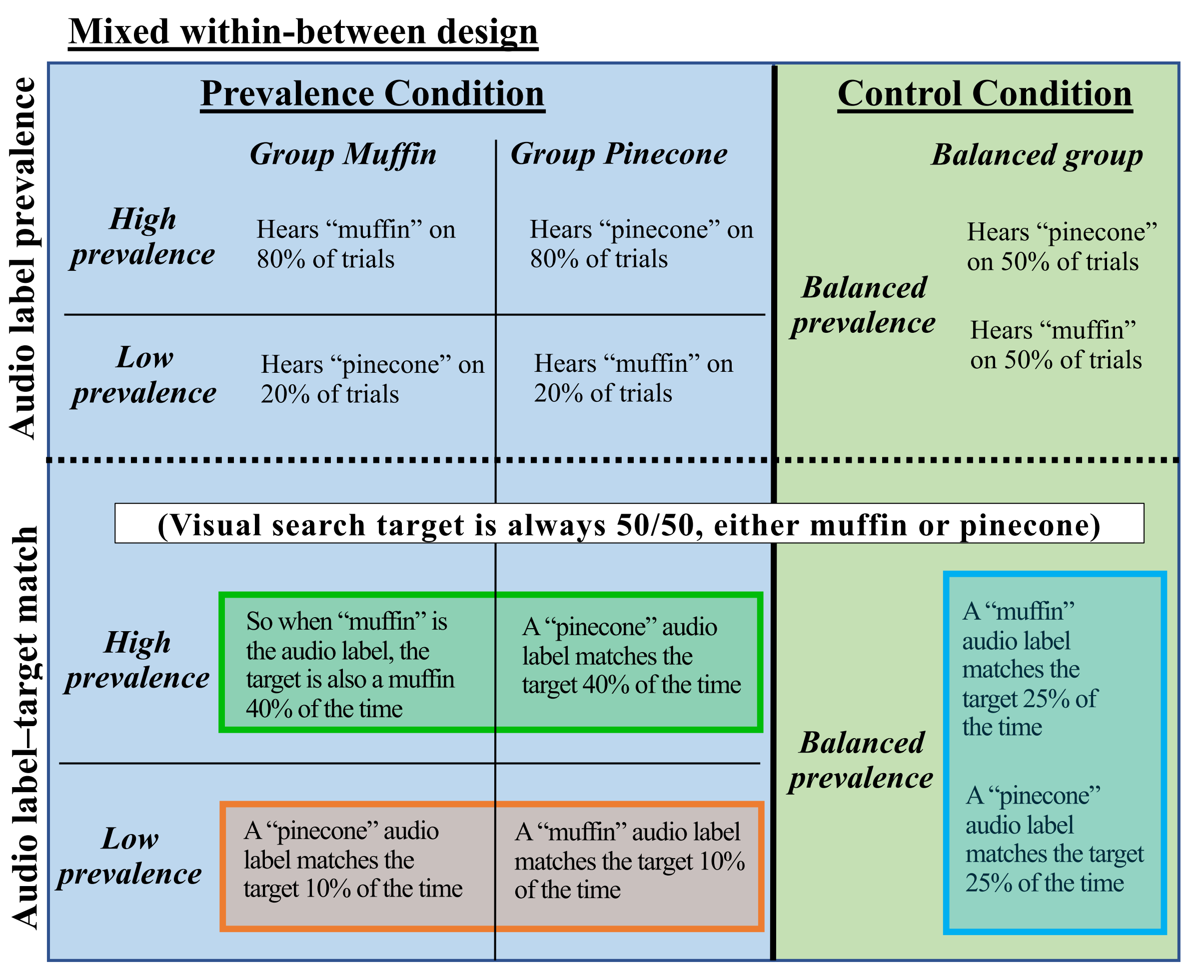

Figure 8. Experiment 3 design. Data from the *high prevalence* and *low prevalence* conditions were respectively grouped together and compared to each other as within-subject factors (see grouped boxes). The *balanced prevalence* control group was compared to the *high* and *low prevalence* conditions as a between-subjects factor.

time (e.g., muffin present on 50% of trials *x* “muffin” heard on 20% of trials = 10% match). In the *control* condition, each target name was heard through headphones on 50% of trials, meaning that each target audio label matched the target image 25% of the time (e.g., muffin present on 50% of trials *x* “muffin” heard on 50% of trials = 25% match). Trial types occurred randomly throughout the entire experiment. There were 200 trials in total, divided between conditions as stated above.

As in previous experiments, search arrays contained one target and 24 distractors. There were approximately 2,000 distractors to select from. The search array was traditional spatial search, with each object placed in a random position on the screen. Visual stimuli were the same as in Experiment 1, and audio stimuli consisted solely of the two target names (i.e., “muffin” and “pinecone”).

Procedure. The procedure for Experiment 3A was very similar to that of previous spatial search experiments (see Figure 1). At the beginning of the experiment, participants were instructed to search for a muffin and a pinecone among distractor objects while listening to audio of the target names. Participants were informed that only one target would be present at a time. Participants were also told that sometimes the audio label would match the target image that was present on the screen, and sometimes it would not. Participants were not explicitly told the prevalence of audio label–target image matches/mismatches. At the start of each trial, there was a “get ready” screen, lasting four seconds, during which the audio label continued to repeat. After the “get ready” display, the search array appeared. Participants were instructed to press the “SPACE” bar when they found a target, as quickly as possible. Audio stopped playing when the spacebar was pressed. RTs were measured from the onset of the search display to the spacebar press. After each response, as in Experiment 1, the search array disappeared and numbers appeared on the screen in locations corresponding to each object for one second. The participants were then given a choice between two numbers from the previous screen, with one representing the correct target location. Participants chose the correct number by pressing the "F" or "J" keys, and "correct" or "incorrect" feedback was displayed. There were ten practice trials at the beginning of the experiment, so that participants were exposed to all different combinations of target image–audio label match vs. mismatch. Experiment 3A lasted approximately one hour, with a break halfway through.

Results

Two (*match:* audio label–target image match vs. no match) *x* 2 (*prevalence:* audio target label high prevalence vs low prevalence) repeated-measures analyses of variance (ANOVAs) were performed for accuracy and RTs. The *control* condition with *balanced prevalence* was analyzed separately and not included in the overall repeated-measures ANOVAs because it was between subjects, but it was included in subsequent planned paired and independent comparisons (*t*-tests). Only correct responses were analyzed for RTs, for which medians were again used. Bonferroni adjustments were applied with a corrected alpha value of .05.

Participants in Experiment 3A were quite accurate across all audio and clarity conditions (overall M = 98.3%). Repeated measures ANOVAs were all insignificant; there was no main effect of *match*, $F(1, 116) = .002$, $p = .965$, partial $\eta^2 = .00$, no main effect of *prevalence*, $F(1, 116) = 2.240$, $p = .137$, partial $\eta^2 = .019$, and no interaction between the two, $F(1, 116) = .311$, $p = .578$, partial $\eta^2 = .003$. Among the planned comparisons, the only differences in accuracy occurred in the *control* condition with *balanced prevalence*; accuracy in trials when the audio target label *matched* the target image (M = 98.5%, SD = 2.7%) was significantly higher than in trials when audio and image *did not match* (M = 97.9%, SD = 3.6%; $t(57) = 2.423$, $p = .019$. Differences in accuracy between all other combinations of *high/low* prevalence and *match/no match* trials were all insignificant, t's < 1.3 (see Figure 9.)

Unlike accuracy, reaction time data revealed many interesting effects. There was a main effect of *match*, $F(1, 116) = 15.607$, $p < .001$, partial $\eta^2 = .119$, wherein RTs were faster when the audio label and target image matched (M = 1,101 ms, SD = 248 ms) than when they did not match (M = 1,156 ms, SD = 291 ms). There was also a main effect of *prevalence*, $F(1, 116) = 4.058$, $p = .046$, partial $\eta^2 = .034$, with slower overall RTs on *high prevalence* trials (M = 1,152 ms, SD = 312 ms) than on *low prevalence* trials (M = 1,105 ms, SD = 265 ms). This effect is driven by the comparatively large difference between *high prevalence* trials that had an audio label–target image *match* (M = 1,107 ms) and had *no match* (M = 1,198 ms), $t(116) = 4.371$, $p < .001$ (see Figure 8). On *low prevalence* trials, there was no such difference between *match* (M = 1,095 ms) and *no match* trials (M = 1,115 ms), $t(116) = 1.118$, $p = .266$. This interaction between *match* and *prevalence* was significant, $F(1, 116) = 7.036$, $p = .009$, partial $\eta^2 = .057$. On *match* trials, performance when hearing the *high prevalence* audio label (M = 1,107 ms) was roughly equivalent compared to when the *low prevalence* audio label was heard (M = 1,095 ms), $t(116) = .446$, $p = .657$. On *no match* trials, however, there was relatively large difference between trials with the *high prevalence* audio label (M = 1,198 ms) and with the *low prevalence* audio label (M = 1,115 ms), $t(116) = 3.091$, $p = .003$. Reasons for this interaction are explored in the *Discussion* section.

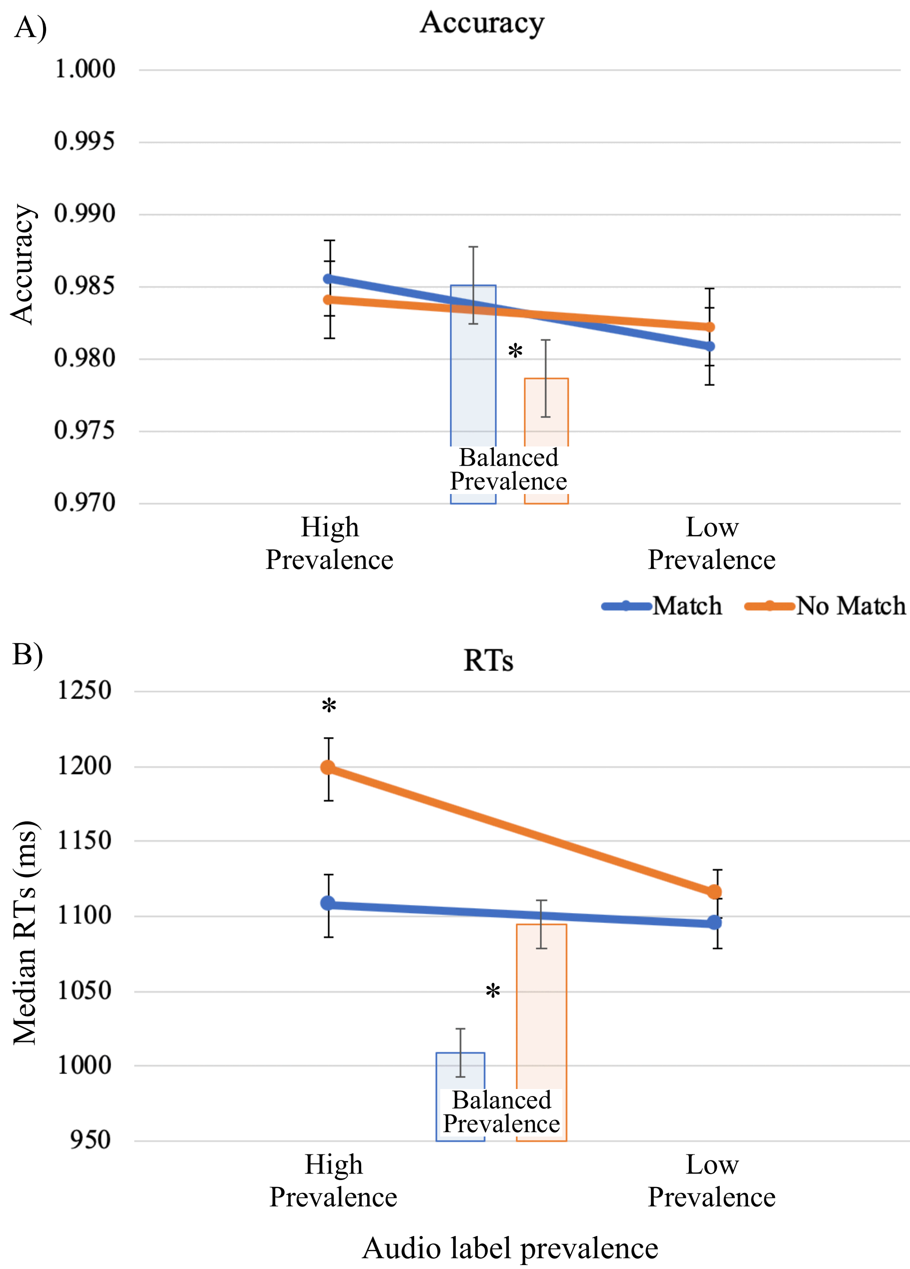

Figure 9. Experiment 3A results, audio label prevalence plotted against audio label–target image match. Line graphs represent data from the *prevalence* condition, and bar graphs represent the *control* condition. Error bars represent ±1 SEM.

Lastly, in the *control* condition with *balanced prevalence*, there was a significant difference between *match* (M = 1,009 ms) and *no match* (M = 1,097 ms) trials, $t(58) = 5.345$, $p < .001$. Independent sample *t*-tests were then calculated to allow comparisons between *high*, *low*, and *balanced* prevalence conditions. When there was an audio label and target image *match*, RTs in the *balanced prevalence* audio label group were significantly faster compared to when the *prevalence* group heard a *high prevalence* audio label, $t(173) = 2.20$, $p = .029$, and compared to when the *prevalence* group heard a *low prevalence* audio label, $t(173) = 1.95$, $p = .050$. When the audio label and target image *did not match*, RTs in the *balanced prevalence* audio label group were nearly significantly faster compared to when the *prevalence* group heard a *high prevalence* audio label, $t(173) = 1.895$, $p = .060$. There was no such difference compared to when the *prevalence* group heard a *low prevalence* audio label, $t(173) = .42$, $p = .676$. This pattern of results is also explored further in the *Discussion* section.

Experiment 3B

Method

Participants. 123 new participants were recruited from the Arizona State University Psychology 101 subject pool, and were given course credit for their participation. Of the 123 participants, 34 were in randomly assigned to *Group Muffin* (participants were not told this), and heard "muffin" on 80% of trials and "pinecone" on 20% of trials. *Group Pinecone* heard "pinecone" on 80% of trials and "muffin" on 20% of trials, and consisted of 29 participants. *Group Muffin* and *Group Pinecone* combined to give the oversampled *prevalence* condition a total of 54 participants. Lastly, the *control* condition consisted of the *balanced group*, which heard "muffin" and "pinecone" each on 50% of trials, and had 60 participants. All 123 participants were native English speakers and had normal or corrected-to-normal vision by self-report. Exclusion criteria was the same as in previous experiments. After these criteria were applied, eight participants were excluded (one from *Group Muffin*, three from *Group Pinecone*, and four from the *balanced group*). A further four participants (one each from *Group Muffin* and *Group Pinecone*, and two from the *balanced group*) were excluded for having an inordinate number of misses, performing well below chance when a target was present, regardless of audio label match or prevalence. This resulted in a total n of 111: 57 in the oversampled *prevalence* condition and 54 in the *control* condition.

Apparatus, stimuli and design. Apparatus, stimuli, and design were identical to that of Experiment 3A, except for the fact that RSVP was used. Each object was displayed in the center of the screen one at a time, each followed by a backwards mask (see experiment 2B *apparatus, stimuli, and design* for description of the mask used). There were 24 distractor objects and one target (either a muffin or a pinecone) on each trial, and the target could not occur as either the first or last object in the display. As in previous experiments, search arrays contained one target and 24 distractors, though the target was absent half the time, in which case it was replaced by an additional distractor so that there were always 25 items in the search array. Distractor stimuli were the same as in previous experiments, and audio stimuli consisted solely of the two target names (i.e., “muffin” and “pinecone”).

Procedure. The RSVP version of Experiment 3 was procedurally very similar to that of the traditional spatial search version (Experiment 3A) and the previous RSVP procedure (Experiment 2B), with relatively minor changes. Participants were instructed to search for a muffin and a pinecone among distractor objects while listening to audio of the target names. Participants were informed that only one target would be present at a time. Participants were also told that sometimes the audio label and target image would match, and sometimes they would not, but participants were not explicitly told the prevalence of matches/mismatches. At the beginning of each trial, there was again a four second "get ready" delay, during which the audio continued to repeat. At the onset of the search array, each of the 25 objects displayed rapidly in the center of the screen, one object at a time. Each item was displayed for 100ms, with a 50ms mask after each image. After the final image disappeared, the audio stopped, and participants were asked to indicate—by pressing the "F" or "J" keys—which of the two targets was present in the array. Accuracy feedback was given after every trial. There were 10 practice trials at the beginning of the experiment. Experiment 3B lasted approximately one hour, with a break halfway through.

Results

As in Experiment 3A, a 2(*match:* audio label–target image match vs. no match) *x* 2(*prevalence:* audio target label high prevalence vs low prevalence) repeated-measures ANOVA was performed for signal detection measures. While descriptives are provided for overall accuracy, further analyses of accuracy would in this case be inappropriate and are not included here. This is because in Experiment 3, *audio label–target image match* is a key variable of interest, but as an RSVP paradigm, Experiment 3B must have target-absent trials. On trials in which the target is absent, there cannot be an audio label–target image *match*, because there is no target image present to match to. This means that comparing true overall accuracy between *match* and *no match* trials is not possible, and instead comparing hit rate is more appropriate. This also results in FAs for *match* and *no match* trials across each *prevalence* condition that are indistinguishable from each other, so FAs are not independently analyzed here[2]. While it is true that FAs are a crucial component in calculating *d'* and *C*, these measure are sufficiently standardized and transformed (see equations 1 – 4 in the *Results* section of Experiment 2B), so that they still carry distinct and important meaning. For this reason, analyses were conducted with *H*, *d'*, and *C* only.

The *control* condition with *balanced prevalence* was again analyzed separately and not included in the overall repeated-measures ANOVAs, but was included in subsequent planned paired and independent comparisons. Bonferroni adjustments were applied on all comparisons, with a corrected alpha value of .05.

Performance in Experiment 3B was high across all audio and clarity conditions ($M_{accuracy}$ = 95.8%). Beginning with just the within-subjects *prevalence* group, there was a main effect of *audio label–target image match* for hit rate, $F(1, 56) = 3.944$, $p = .051$, partial $\eta^2 = .065$, which revealed that overall hit rate was higher on *match* trials (M_H = .965, SD_H = .062) than on *no match* trials (M_H = .930, SD_H = .143; see Figure 10). This effect of *match* on hit rate was larger when hearing the *high prevalence* audio label (M_{match} = .968, M_{no_match} = .928, $t(56) = 2.279$, $p = .026$) than when hearing the *low prevalence* audio label (M_{match} = .962, M_{no_match} = .933, $t(56) = 1.431$, $p = .158$; See Figure 10). However, there was no main effect of *audio label prevalence* and no interaction for hit rate, *F*'s < .6.

For sensitivity (*d'*), there was a main effect of *match*, $F(1, 56) = 5.898$, $p = .018$, partial $\eta^2 = .095$, which showed that participants were on average more sensitive on

[2] It should nevertheless be noted that false alarm rates on trials containing the *high prevalence* audio label (M = .024) did not differ from FAs on *low prevalence* trials (M = .026), $t(56) = .445$, $p = .658$. This is the only comparison of false alarms that can be made.

match trials ($M_{d'}$ = 2.577, $SD_{d'}$ = .468) than on *no match* trials ($M_{d'}$ = 2.46, $SD_{d'}$ = .554). When hearing the *high prevalence* audio label, sensitivity was higher on *audio label–target image match* trials (M = 2.846) than on *no match* trials (M = 2.683), $t(56) = 2.863$, $p = .006$. When hearing the *low prevalence* audio label, however, sensitivity did not significantly vary between *match* (M = 2.309) and *no match* (M = 2.236) trials, $t(56) = 1.402$, $p = .167$. This interaction between *match* and *prevalence* for *d'* was very nearly significant, $F(1, 56) = 3.333$, $p = .072$, partial $\eta^2 = .056$. There was also a very large main effect of *prevalence* for sensitivity, $F(1, 56) = 127.509$, $p < .001$, partial $\eta^2 = .694$, which indicated that participants were on average much more sensitive when hearing the *high prevalence* audio label ($M_{d'}$ = 2.765, $SD_{d'}$ = .480) than when hearing the *low prevalence* audio label ($M_{d'}$ = 2.272, $SD_{d'}$ = .304). On *match* trials, sensitivity was much higher when hearing the *high prevalence* audio label (M = 2.846) than when hearing the *low prevalence* label (M = 2.309), $t(56) = 11.520$, $p < .001$. There was also a very large difference in sensitivity on *no match* trials, where participants were much more sensitive on *high prevalence* audio trials (M = 2.683) than on *low prevalence* trials (M = 2.236), $t(56) = 8.381$, $p < .001$.

When examining response bias, there was a main effect of *match* for measures of *C,* $F(1, 56) = 5.90$, $p = .018$, partial $\eta^2 = .095$; participants were on average less conservative (indicated by a smaller *C*) on *match* trials (M_C = .118, SD_C = .195) than on *no match* trials (M_C = .201, SD_C = .305). This effect was driven by the significant difference in bias between *match* (M = .085) and *no match* trials (M = .200) when hearing the *high prevalence* audio label, $t(56) = 2.865$, $p = .006$. On *low prevalence* trials, there is no such difference between *match* (M = .151) and *no match* trials (M = .202), $t(56)$ =

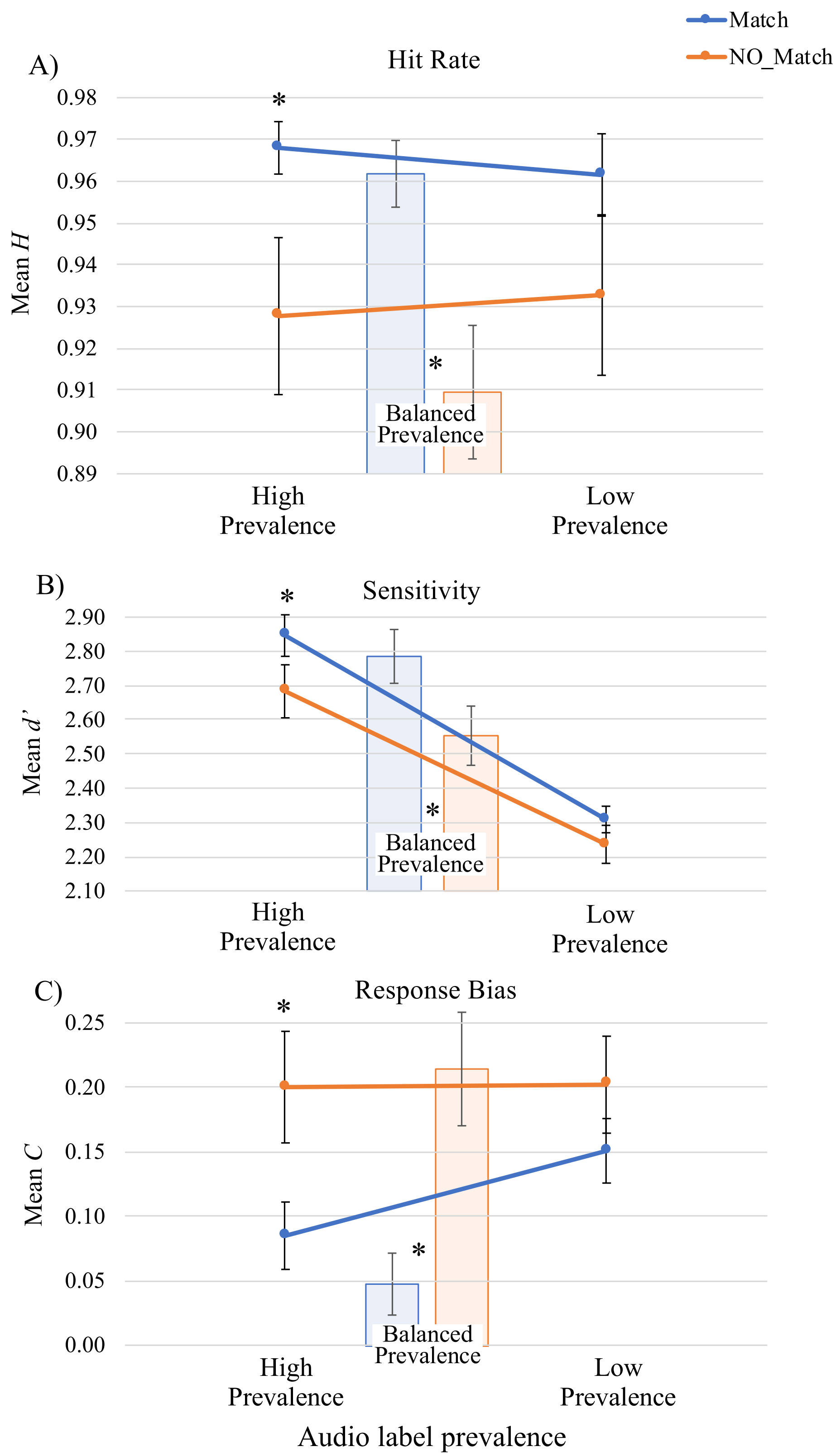

Figure 10. Experiment 3B data for hit rate (A), sensitivity (B), and bias (C). Line graphs represent data from the *prevalence* condition, and bar graphs represent the *control* condition. Error bars represent ±1 SEM.

1.399, $p = .167$. There was no main effect of *prevalence* for bias, $F(1, 56) = 1.19$, $p = .280$, partial $\eta^2 = .020$, though there was a marginally significant difference between *high* ($M = .085$) and *low prevalence* ($M = .151$) audio labels on *match* trials, $t(56) = 1.942$, $p = .057$ (there was no such difference on *no match* trials, $t = .051$). There was a marginally significant interaction between *match* and *prevalence*, $F(1, 56) = 3.351$, $p = .071$, partial $\eta^2 = .057$ (see Figure 10).

Lastly, for the *balanced prevalence* control group, paired sample comparisons revealed that performance was overall better when the audio label *matched* the target image. Hit rate was significantly higher for *match* ($M = .962$) than *no match* ($M = .910$) trials, $t(53) = 3.568$, $p = .001$. As in the *prevalence* group, participants in the *control* group were more sensitive on *match* trials ($M_{d'} = 2.797$) than *no match* trials ($M_{d'} = 2.561$), $t(53) = 4.559$, $p < .001$. Participants were also less conservative on *match* trials ($M_C = .048$) than on *no match* trials ($M_C = .215$), $t(53) = 4.560$, $p < .001$. Independent sample *t*-tests were then used to compare hit rate, sensitivity, and bias between the *prevalence* and *control* groups. There were no significant differences between the *high prevalence* and *balanced prevalence* conditions across all *match* and *no match* conditions and all signal detection measures, all *t*'s < 1.17. Between the *low prevalence* and *balanced prevalence* conditions, on *match* trials sensitivity was significantly lower in the *low prevalence* condition ($M_{d'} = 2.309$) than in the *balanced prevalence* condition ($M_{d'} = 2.797$), $t(109) = 5.559$, $p < .001$. This was also true for *no match* trials, where sensitivity was lower in the *low prevalence* condition ($M_{d'} = 2.236$) than in the *balanced prevalence* condition ($M_{d'} = 2.561$), $t(109) = 3.163$, $p = .002$. Participants were also more conservative in the *low prevalence* condition ($M_C = .151$) than in the *balanced prevalence* condition (M_C = .048) for *match* trials, $t(109) = 2.946$, $p = .004$. This difference in bias was not present for *no match* trials, nor were there any differences present between *low* and *balanced prevalence* conditions for hit rates or false alarm rates, *t*'s < .92.

Discussion

Experiment 3 investigated whether the relative prevalence of an audio label might modulate the label feedback effect, and utilized both traditional spatial visual search (Experiment 3A) and RSVP (Experiment 3B). These experiments drew inspiration from the *low-prevalence effect* (LPE), a phenomenon wherein observers are exceedingly more likely to miss targets that occur rarely relative to the same targets that occur frequently (Wolfe et al., 2005). In the current experiment, participants searched for the same two targets (a muffin and a pinecone) throughout the entire experiment. In the *prevalence* condition, subjects heard one of the two target names played through headphones on 80% of trials (*high prevalence*), and the second of the two target names was heard 20% of the time (*low prevalence*). In the *control* condition, subjects heard each target name 50% of the time (*balanced prevalence*). I expected that there would be an overall effect of *match* across conditions, where performance would be better when the target image and the target audio label match, compared to trials in which there was *no match*. I predicted that this effect would be strongest on trials where the participants heard the *high prevalence* audio label, compared to on trials with the *low prevalence* audio label.

For the most part, results supported my hypotheses. Across both Experiments 3A and 3B, there was evidence of a label feedback effect, which manifested as either a main effect of *match* or in a *match x prevalence* interaction. In spatial search (Experiment 3A), when hearing the *high prevalence* audio label, RTs were significantly faster when the label matched the target image compared to when there was a mismatch. This difference was not significant when hearing the *low prevalence* audio, however, and there was no difference between *high* and *low prevalence* trials when there was an audio label–target image match. For both the *prevalence* group and the *control* group, the audio label matched the target image 50% of the time; what varied was *which* audio label was more prevalent, but given that a target was always present on 50% of trials, the actual predictive values of each audio label were equivalent. Presumably what is driving the effect, then, is target template integrity. There was no prevalence effect when there was an audio–target *match* (i.e., no label feedback *facilitation*), but on the relatively rare occasion that the *high prevalence* audio label did *not* match the target, performance suffered; the target template was not as salient. This pattern of effects offers support for the argument that a main mechanism of the label feedback effect is the *protection* of target template integrity, rather than facilitation of visual processing through lexical activation. The frequency of the *high prevalence* target label preserves the template of the corresponding target image, and subsequently the target template of the *low prevalence* target is more susceptible to degradation. This account is consistent with previous low prevalence research (e.g., Hon & Tan, 2013; Hout et al. 2015; Wolfe et al., 2005) that demonstrates that *low prevalence* stimuli are perceived differently than *high prevalence* stimuli.[3]

One interesting finding is that the magnitude of the label feedback effect in the *balanced prevalence* control group is often somewhat larger than in the *prevalence* group. There is a label feedback effect for accuracy—which, even though performance is

[3] It's doubtful that the audio stimulus is missed altogether, as is often the case for low-prevalence visual stimuli in a search array, even when they are directly fixated (Hout, Walenchok, Goldinger, & Wolfe, 2015. The sudden change of the only auditory stimulus from the *high-prevalence* audio label to the *low-prevalence* audio label would likely not go unnoticed. However, future studies could utilize pupillometry to observe if a physiological response indicates whether the stimulus change is processed.

arguably at ceiling, was not present for the *prevalence* group—and RT effect sizes are larger (see Figure 9). This disparity is also present in signal detection measures in RSVP (see Figure 10). This difference is likely a product of the experimental design; in the *balanced prevalence* control condition, there was no prevalence to learn—each audio label matched the target image 50% of the time, so there was no extra level of "protection" or "vulnerability" associated with either of the target templates in working memory. In other words, what is observed in the *balanced prevalence* control condition is purely a label feedback effect. For the *prevalence* groups, the *high* and *low prevalence* of audio labels changed the perception of those labels, and the result is an interaction of the label feedback effect and the *low prevalence effect*.

A degree of this interaction can possibly be explained by adding priming to the conversation. Figure 11 proposes a conceivable explanation of reaction time results (from Experiment 3A) by examining varying degrees of priming and target template integrity on muffin-target trials (the same concept would of course apply to the pinecone target, as well). *Group Muffin* heard "muffin" on 80% of trials, and so had strong muffin-target template integrity. *Group Pinecone* heard "muffin" on only 20% of trials, and so had relatively weak muffin-target template integrity. For the *balanced group*, target template integrity was presumably equivalent for both the muffin and the pinecone, as each target name was heard on 50% of trials. First, consider trials where the muffin was the present target image *and* the audio label. *Group Muffin* had strong muffin template integrity *and* the concept of "muffin" was now primed through the audio stimulus, resulting in very short search RTs (938 ms). On these trials *Group Pinecone*

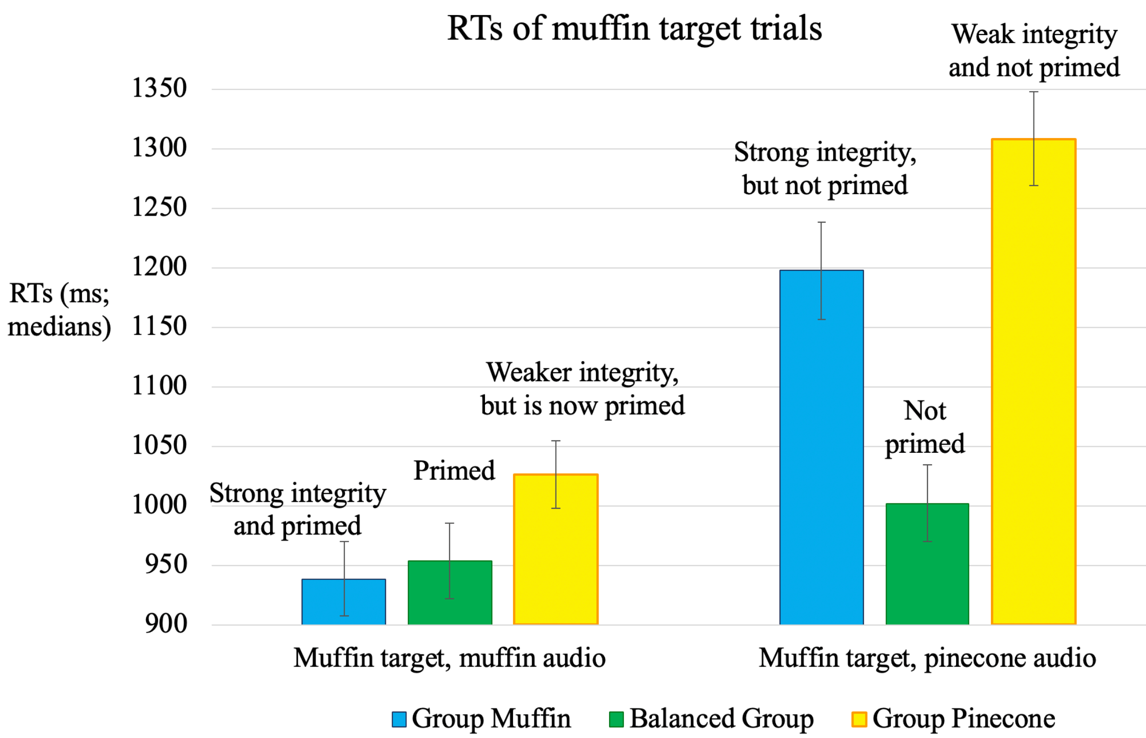

Figure 11. Target template integrity and priming in the label feedback and low prevalence effects through muffin-target reaction time data (Experiment 3A). Left: Trials with the muffin target image present on-screen *and* the "muffin" audio label. Right: Muffin-target trials with the "pinecone" audio label. *Group Muffin* (blue) heard "muffin" on 80% of trials; *Group Pinecone* (yellow) heard "muffin" on 20% of trials; *Balanced Group* (green) heard "muffin" on 50% of trials. Error bars represent ±1 SEM.

had relatively weak muffin template integrity, but the concept of muffin was now primed, so while RTs were slower than that of *Group Muffin*, they were still relatively fast (1,026 ms). The *balanced group* had an unaffected muffin target template, but muffin was now primed, and so RTs were somewhere in the middle (954 ms). Next, consider trials where muffin was the present target image, but the audio label was "pinecone". *Group Muffin* had strong muffin template integrity, but a competing concept was primed instead. This relatively rare occurrence (10% of all trials) is jarring, and results in a sizeable 260 ms jump in RTs (1,197 ms). *Group Pinecone* had weak muffin template integrity, and not only was the concept of “muffin” not primed on these trials, but *Group Pinecone*’s *high prevalence* audio label (i.e., “pinecone”) *was* primed, leaving them very perceptually unprepared to detect the muffin image (M_{RTs} = 1,308 ms). The *balanced group* loses the priming benefit on these trials, but there is no difference in target template integrity, so the effect is not nearly as detrimental (M_{RTs} = 1,002 ms). It is presumably through this combination of priming (facilitation) and target template maintenance (integrity/protection) that the label feedback effect and the low prevalence effect function.

It is difficult to determine, however, the exact degree to which each effect influences these results, and the specific mechanisms that belong to each effect. Conceptually, the *balanced prevalence* control group reflects only the label feedback effect, and in that sense it can be argued that priming is indeed an important component of the label feedback effect. In the *prevalence* group, on the other hand, the relative prevalence of a target’s audio label very strongly impacted search performance, and therefore likely plays a role in a target’s template integrity to at least some extent. We see from Experiment 3B that the *low prevalence* audio label resulted in more conservative response bias and poorer sensitivity. This substantiates previous research, which suggests that poorer performance with *low prevalence* stimuli arises from a failure of perception (Hout et al., 2015). Even though the present study is the first to examine the label feedback effect with signal detection measures specifically in visual search tasks, results demonstrated that label feedback (i.e., *match* trials) increased sensitivity and reduced the conservative response bias, which is consistent with previous findings (Lupyan & Ward, 2013).

One limitation of the current paradigm was that there was a relatively small number of *low-prevalence* trials—20% of all trials, compared to 80% (*high prevalence*) and 50% (*balanced prevalence*)—but even though that condition has slightly higher variance, the effects are fairly robust. Nevertheless, a longer experiment with additional trials might yield more reliable results. One other item to be improved upon in future experiments is the choice of target stimuli. Despite the fact that they were each sized to 100 x 100 pixels and even had similar coloring, the muffin target image was for some reason easier to find than the pinecone target, and was on average found 100-200 ms faster. Of course, counterbalancing between which target was *high prevalence* and which was *low prevalence* means that the statistical comparisons between experimental conditions are still valid. However, different target stimuli with more closely-matched processing times could potentially yield cleaner data.

GENERAL DISCUSSION

The present experiments were designed to explore the boundaries and mechanisms behind the label feedback effect through a series of spatial and RSVP visual search tasks. Experiment 1 extended the effect beyond self-directed speech by replicating the work of Hebert et al. (under review) with audio stimuli played through headphones, and found that listening produces the label feedback effect. Experiment 2 examined the label feedback effect under conditions of varying object clarity by manipulating the level of blur in search images, and results showed that hearing target names improved performance, even (and sometimes especially) when conditions were difficult or noisy. Finally, Experiment 3 investigated the interaction between the label feedback effect and the low prevalence effect by manipulating the relative prevalence of a target's audio label during dual-target search, and found that the two effects combined in interesting ways to each impact a target's perception.

The findings presented here substantiate previous literature on the label feedback and low prevalence effects, and offer new insights. To the author's knowledge, no research has examined effects of prevalence in any modality other than visual. Common real-world examples of the low prevalence effect include airport baggage screeners searching images for contraband, or radiologists searching x-rays and CAT scans for malignancies. In these cases, even expert observers still miss rare target items upwards of 30% of the time (Evans et al., 2013; Evans et al., 2011; Reed, Ryan, McEntee, Evanoff, & Brennan, 2011; Wolfe et al., 2013), often even when they directly fixate the rare item (Hout et al., 2015). A single auditory stimulus is perceptually very different from an entire array of visual stimuli, however, and so the change from one very common sound to a rare sound is much less likely to go unnoticed. Even when accuracy remains high—as it did here—infrequently-occurring targets are processed much slower than high-prevalence targets (Laberge & Tweedy, 1964; Miller & Pachella, 1973), and low-prevalence targets are more attentionally demanding (Hon & Tan, 2013). So even though lower-prevalence auditory stimuli are likely actively attended, they nevertheless come with a significant perceptual cost, substantiating the notion that issues of prevalence are complex and persistent, and suggesting that they are likely not limited to the visual world.

Might label feedback mitigate some of these perceptual costs? Previous research has found that label feedback increases target sensitivity (*d'*); for example, target labels can "boost" images near perceptual threshold into visual awareness (Lupyan & Ward, 2013). Across both RSVP experiments in the current study, hearing the name of the target image indeed increased sensitivity to the target, and additionally resulted in a small criterion shift—an effect not previously reported in the literature—wherein response bias (*C*) became less conservative. This was the case even when visual conditions were noisy or difficult (as in *blur* conditions in Experiment 2B), and when a stimulus occurred infrequently (Experiment 3B). While these effects were often smaller for targets associated with a low-prevalence audio label (relative to those with a high-prevalence label), they were present nevertheless. This could have potential real-world implications—it suggests, for example, that baggage-screeners might improve sensitivity and hit rates for contraband by continuously processing weapon names. Research suggests that prevalence costs increase in multiple-target search, however, (Godwin, Menneer, Cave, & Donnelly, 2010; Menneer, Donnelly, Godwin, & Cave, 2010), so additional research is needed to determine at what point this strategy might no longer be helpful or practical.

As for the fate of the label feedback effect, the findings presented here provide additional evidence that the effect is driven by working memory and attentional processes, rather than facilitation of visual processing through lexical activation. The current findings suggest that the key mechanisms behind the label feedback effect are priming and target template maintenance. An important distinction here is that not only can label feedback facilitate performance to a degree (i.e., through priming), the ability to *protect* against distraction and other performance-inhibiting factors in favor of maintaining target template integrity is equally, if not more, important. Visual search creates natural challenges for template integrity in working memory: Theoretically, as search proceeds, every fixated object is analyzed—Its visual features are perceived, and its identity may be appreciated (along with its name; Meyer, Belke, Telling & Humphreys, 2007; Walenchok, Hout & Goldinger, 2016). In RSVP especially, this all takes place exceptionally quickly as the observer is forced to analyze each and every object in rapid succession. As the perceptual system serially evaluates each object one after another, it creates natural interference with a search target's mental template. Label feedback—in this case, hearing a target's name—repeatedly activates the search template, which strengthens its integrity (perhaps increasingly so over time) and defends against interference.

However, as the discrepancy between a target's image and its template increases (as in *full blur* stimuli), label feedback becomes increasingly unhelpful. There seems to be a "sweet spot" (potentially in *minimal blur* stimuli, for example,) where stimuli are similar enough to their templates that hearing a stimulus' name still facilitates search, but visual conditions are *just difficult enough* that without label feedback, appreciating the target object is much more difficult. (Future paradigms could utilize a more continuous spectrum of varying object clarity to pinpoint where this shift takes place.) Along those lines, when a target's label occurs with high prevalence, that target's template is better preserved and more readily available. Conversely, the templates of low-prevalence targets have relatively weaker integrity and are more susceptible to degradation. When priming on a trial is added to the equation through label feedback, it can modulate prevalence effects created by relative target frequency (or vice versa). While the degree to which priming and target template integrity each independently contribute to effects of label feedback and prevalence is not yet clear, it is apparent from these findings that they do in fact interact in a significant way.

REFERENCES

Brady, T. F., Konkle, T., Alvarez, G. A. and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, USA*, 105, 14325-14329. https://doi.org/10.1073/pnas.0803390105

Brophy, A. L. (1986). Alternatives to a table of criterion values in signal detection theory. *Behavior Research Methods, Instruments, & Computers, 18,* 285-286. https://doi.org/10.3758/BF03204400

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Erlbaum.

Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PLoS ONE, 8,* e64366. http://dx.doi.org/10.1371/ journal.pone.0064366

Evans, K. K., Tambouret, R. H., Evered, A., Wilbur, D. C., & Wolfe, J. M. (2011). Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. *Archives of Pathology & Laboratory Medicine, 135,* 1557–1560. http://dx.doi.org/10.5858/arpa.2010- 0739-OA

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. https://doi.org/10.3758/BF03193146

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00335

Felsten, G., & Wasserman, G. S. (1980). Visual masking: Mechanisms and theories. *Psychological Bulletin*, *88*(2), 329.

Godwin, H.J., Menneer, T., Cave, K.R., Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition, 18:*1439–1463. http://dx.doi.org/10.1080/13506285.2010.500605.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Hebert, K.P., Goldinger, S.D., & Walenchok, S. C. (under review) Eye Movements and the Label feedback effect: Speaking Modulates Visual Search via Template Integrity. *Journal of Cognition.*

Hon, N., & Tan, C. H. (2013). Why rare targets are slow: Evidence that the target probability effect has an attentional locus. *Attention, Perception, & Psychophysics, 75*(3), 388-393.

Hout, M.C., and Goldinger, S.D. (2010) Learning in repeated visual search. *Attention, Perception, & Psychophysics, 72*:1267-1282. http://dx.doi.org/10.3758/APP.72.5.1267.

Hout, M.C., Goldinger, S.D. (2012) Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 38*:90–112. http://dx.doi.org/10.1037/a0023894.

Hout, M.C. & Goldinger, S.D. (2015) Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*: 128. https://doi.org/10.3758/s13414-014-0764-6

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the *Massive Memory* picture database. *PLoS ONE, 9*: e112644. doi: 10.1371/journal.pone.0112644.

Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 977.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*, 558-78.

Kunar, M. A., Rich, A. N., & Wolfe, J. M. (2010). Spatial and temporal separation fails to counteract the effects of low prevalence in visual search. *Visual Cognition*, *18*(6), 881–897. http://doi.org/10.1080/13506280903361988

Laberge, D., & Tweedy, J. R. (1964). Presentation probability and choice time. *Journal of Experimental Psychology, 68,* 477–481.

Lupyan, G. (2007). *The label feedback hypothesis: Linguistic influences on visual processing* (unpublished PhD thesis). Pittsburgh, PA: Carnegie Mellon University.

Lupyan G. (2012) Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

Lupyan, G., and Swingley, D. (2012). Self-directed speech affect visual search performance. *Quarterly Journal of Experimental Psychology*, i*First*, 1-18. https://doi.org/10.1080/17470218.2011.647039

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196-14201.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. 2nd Edition. Psychology press.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185.

McGill, W.J., Gibbon, J. (1963). The general-gamma distribution and reaction times. *Social Sciences*, *1963*.

McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.

Menneer, T., Donnelly, N., Godwin, H.J., Cave, K.R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied, 16*:133–144. http://dx.doi.org/10.1037/a0019569. [PubMed: 20565198]

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*, 710–716. https://doi.org/10.3758/BF03196826

Ratcliff, R., Murdock, B.B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *1976*(190—214).

Reed, W.M., Ryan, J.T., McEntee, M.F., Evanoff, M.G., Brennan, P.C. (2011) The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology, 258*:938–943. doi: radiol.10101090. [PubMed: 21248231]

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: The role of stable and transient states of the observer. *Frontiers in Human Neuroscience, 8*, 18.

Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, *85*(2), 198.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149.

Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics, 78,* 2633-2654.

Williams, C.C (2010) Not all visual memories are created equal. *Visual Cognition, 18*:2, 201-228, DOI: 10.1080/13506280802664482

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision, 13*(3):33, 1–9, http://www.journalofvision.org/content/13/3/33, doi:10.1167/13.3.33.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: Rare items often missed in visual searches. *Nature*, *435*(7041), 439-440. https://doi.org/10.1038/435439a

APPENDIX A

PYTHON CODE FOR CREATION OF AUDIO STIMULI

```python
import csv

#read the list of words to convert to audio

with open('targ_nonw.csv', 'r') as f:
   reader = csv.reader(f)
   word_list = list(reader)

print(word_list)



from gtts import gTTS

import os

# gtts = google text to speech
# The text that you want to convert to audio

for word in word_list:
    for x in word:
        #print (x)

            mytext = (x)
            # Language in which you want to convert
            language = 'en'

            # Passing the text and language to the engine,
            # slow=True tells the module that the converted audio
            # should have a normal speaking speed

            myobj = gTTS(text=mytext, lang=language, slow=True)

            # Saving the converted audio in a mp3 file
            # named with the word root + .mp3

            myobj.save((mytext) + ".mp3")
```

APPENDIX B

PYTHON CODE FOR CREATION OF BLURRED STIMULI

```python
from PIL import Image
from PIL import ImageFilter
import glob, os, fileinput, sys

##--------------MINIMAL BLUR (BLUR RADIUS = 1)--------------##
#for every item in X folder that ends in X,
#apply a basic blur to the image
for entry in os.scandir('/Users/katehebert/Dropbox (ASU)/DISSERTATION/P
rogramming/LFE_Audio/Program/Resources'):
    if entry.path.endswith('_blur.bmp'):
        continue
    if entry.path.endswith('.bmp'):
        img = Image.open(entry.path)
        img = img.filter(ImageFilter.BoxBlur(1))
        #img.show() optional, to verify that it worked

        #and then resave each of those new images under a new filename
        #Split the original filename into name and extension
        (name, extension) = os.path.splitext(entry.path)
        #Save with "_blur" added to the filename
        img.save(name + '_blur' + extension)


##-------REPEAT PROCESS WITH FULL BLUR (BLUR RADIUS = 3)-------##

for entry in os.scandir('/Users/katehebert/Dropbox (ASU)/DISSERTATION/P
rogramming/LFE_Audio/Program/Resources'):
    if entry.path.endswith('_blur.bmp'):
        continue
    if entry.path.endswith('.bmp'):
        img = Image.open(entry.path)
        img = img.filter(ImageFilter.BoxBlur(3))
        #img.show() optional, to verify that it worked
        (name, extension) = os.path.splitext(entry.path)
        img.save(name + 'full_blur' + extension)
```

APPENDIX C

IRB APPROVAL DOCUMENT

EXEMPTION GRANTED

Stephen Goldinger
Psychology
480/965-0127
goldinger@asu.edu

Dear Stephen Goldinger:

On 1/23/2015 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Language and Perception Research
Investigator:	Stephen Goldinger
IRB ID:	STUDY00002135
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• HRP-503a - Language and Perception Research PROTOCOL SOCIAL BEHAVIORAL-3.docx, Category: IRB Protocol; • Language and Perception _ Eye Tracker Explanation.pdf, Category: Participant materials (specific directions for them); • Language and Perceptiopn Recruitment Script-2.pdf, Category: Recruitment Materials; • LanguageAndPerception_coverletter-3.pdf, Category: Consent Form;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 1/23/2015.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator
cc: Katherine Jones
Stephen Goldinger