Journal of Memory and Language 84 (2015) 37–48

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production

Cassandra L. Jacobs *, Loretta K. Yiu, Duane G. Watson, Gary S. Dell

University of Illinois at Urbana-Champaign, USA

ARTICLE INFO

Article history:
Received 5 September 2014
revision received 13 May 2015

Keywords:
Repetition priming
Duration
Inner speech
Homophone production

ABSTRACT

Acoustic reduction for repeated words could be the result of articulation and motor practice (Lam & Watson, 2014), facilitated production (Gahl, Yao, & Johnson, 2012; Kahn & Arnold, 2015), or audience design and shared common ground (Galati & Brennan, 2010). We sought to narrow down what kind of facilitation leads to repetition reduction. Repetition could, in principle, facilitate production on a conceptual, lexical, phonological, articulatory, or acoustic level (Kahn & Arnold, 2015). We compared the durations of the second utterance of a target word when the initial production was aloud or silent. The silent presentation either involved unmouthed or mouthed inner speech. Overt production, unmouthed and mouthed inner speech all led to reduction in target word onsets, but target word durations were only shortened when a word was initially said aloud. In an additional experiment, we found that prior naming of a homophone of the target word also led to duration reduction. The results suggest that repetition reduction occurs when there is a recently experienced auditory memory of the item. We propose that duration may be controlled in part by auditory feedback during production, the use of which can be primed by recent auditory experience.

Introduction

Acoustic reduction is a change in the speech signal that makes an utterance, or part of it, less prominent. Reduction is multidimensional: reduced words are shorter in duration, more co-articulated, quieter, and have flatter pitch contours (Bard et al., 2000; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fowler & Housum, 1987). Words that are acoustically less prominent are often predictable (e.g. Aylett & Turk, 2004; Bell et al., 2009), in common ground, and given (Bard et al., 2000). Repeated and predictable words may be reduced because they are easy to produce by speakers (Gahl, Yao, & Johnson, 2012) and easy for listeners to understand (Lieberman, 1963). One claim is that reduction is an emergent property of the production system; speakers reduce when production is facilitated or easy. This paper explores the production components that contribute to reduction that is associated with repeated words, which we call *repetition reduction*. If a speaker says, "I saw a *clown* the other day. The *clown* was unhappy," the second utterance of clown will typically be reduced.

One hypothesis proposed by Kahn and Arnold (2012, 2015) is that priming at any level of production leads to reduction, including repetition reduction. Under the *facilitation reduction hypothesis*, facilitation can arise from priming if it occurs at the conceptual, syntactic, lexical, phonological, or articulatory level. In a referential communication task, Kahn and Arnold (2012) found that referring expressions were reduced both when a speaker could anticipate the target to be named from the affordances of the task and when that referent had been previously

* Corresponding author at: Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, USA.

E-mail address: cljacob2@illinois.edu (C.L. Jacobs).

http://dx.doi.org/10.1016/j.jml.2015.05.004
0749-596X/

mentioned, suggesting that priming at conceptual and lexical levels can lead to reduction. Similarly, Bell et al. (2009) found a correlation between word duration and predictability, frequency, and repetition. They argued that duration reduction is linked to the ease of lexical retrieval.

An alternative hypothesis is that effects of facilitation on repetition reduction are limited to just the phonological representations that support articulation. Some evidence from the literature supports this hypothesis. In a referential communication task, Lam and Watson (2010) found that words that were expected were produced with less intensity than words that were unexpected, but words that were repeated were produced with shorter durations, suggesting that the phonological priming that arises from repetition is more strongly linked to duration than the cognitive factors that underlie speaker expectations. In another study, Lam and Watson (2014) investigated whether the givenness of a referent or the givenness of a name were critical for repetition reduction. In a picture description task, they found that reduction only occurred when a referring expression was repeated, independent of whether the referent was new or given. Furthermore, repetition of a referent by itself did not lead to reduction. Similarly, Kahn and Arnold (2015) found that simply hearing a word on the first trial was sufficient to create reduction when that word was then said aloud (see also Bard et al., 2000). These results, like the previous ones, suggest that reduction has some basis in the priming of the form of the word, and that word form may have a privileged effect on word duration.

Additional evidence for the form-facilitation hypothesis comes from work by Kahn and Arnold (2015). In their study, participants described two events in which objects undergo simple actions (e.g. "The cat shrinks", "The barn flashes", or "The artichoke spins"). In *congruent* conditions, the noun in the second sentence (e.g. "cat", "barn") was mentioned in the first. In the *incongruent* condition, the noun in the second utterance was new. Critically, Kahn and Arnold also conducted a study in which the first utterance was just a single noun. They manipulated whether that noun was spoken aloud or produced as inner speech. The second utterances were entire sentences and were always aloud. This inner speech manipulation allowed Kahn and Arnold to probe the level of production at which facilitation occurs and which levels affect reduction. Inner speech lacks sound and articulation, and has been argued to be phonologically impoverished. Nonetheless, it is conceptually and even lexically just as rich as overt speech (Oppenheim & Dell, 2008). Finding robust repetition reduction in duration when the first utterance is produced as inner speech would suggest that priming lower-level linguistic, articulatory, and auditory representations is not necessary for reduction.

Kahn and Arnold (2015) found that inner speech and overt speech led to speeded production in congruent conditions, most dramatically for the time to start the sentence, thus demonstrating that priming at higher levels in the production system can speed the time to articulation. For the actual duration of the second target word, Kahn and Arnold found that congruent targets were 22 ms shorter than incongruent ones in aloud trials, while there was only a 2 ms shortening in inner speech trials. Because repetition reduction only occurred in the aloud condition and not in the inner speech condition, these data suggest that reduction depends on engaging form-based representations that accompany overt articulation, including low-level phonological, articulatory, or auditory representations.

In the present study, we examine more fully the extent to which repetition reduction is the result of articulation, as opposed to other mechanisms. First, we sought to replicate the results of Kahn and Arnold's (2015) inner speech study in a more highly powered experiment. The goal was to determine whether there was any effect of facilitation on reduction in an inner speech condition. The presence of reduction in an inner speech condition would suggest that priming lexical or message level representations in production can potentially lead to repetition reduction. If priming only occurs in overt speech, this would suggest that only priming lower-level, form-based representations affects the duration of words.

In the first two experiments, we examined whether unmouthed (Experiment 1) and mouthed (Experiment 2) inner speech led to duration reduction. Speakers described two events, each containing a noun that could be either *repeated* (referred to as congruent in Kahn and Arnold) or *not repeated* (incongruent). Both types of inner speech are silent, but mouthed inner speech engages articulatory processes. Oppenheim and Dell (2008, 2010) found that mental speech errors made during mouthed inner speech showed the classic phonetic similarity effect – similar consonants were more likely to exchange with each other – but this effect was diminished in unmouthed inner speech. Unmouthed inner speech thus appears to lack full access to the phonological details of a word, while this information is available to mouthed inner speech. If articulatory practice contributes to reduction, we should expect repetition reduction with mouthed inner speech, but not unmouthed inner speech. If repetition reduction is linked to facilitation at higher levels of the production system, we expect repetition reduction in both mouthed and unmouthed inner speech.

Experiment 1

Method

Participants

Participants were 55 native speakers of English who were recruited through the University of Illinois Paid Subject Pool. Each person received $8 for their participation in the study.

Materials

A total of 96 colorized images from the Snodgrass and Vanderwart (1980) dataset (Rossion & Pourtois, 2004) were used, following Lam and Watson (2010). 48 of these images were designated as the critical items. From these critical items, 24 pairs were created (e.g. pear-ruler, axe-rooster). There was no phonological overlap between the names of pair members.

Design

On each trial, participants were presented with an array of four objects. Then two events occurred. First, one object (the prime) shrank, and then another object (the target) flashed. The participant described both events. On half of the critical trials, the prime and the target were the same referent (repeated condition), and in the other half, they were different (not repeated). Whether the prime was produced aloud or not was also manipulated. In half of the critical trials, the prime event was said as unmouthed inner speech, and in the other half of the critical trials the primes were said as overt speech. In all cases, the target events were described overtly.

The experiment was separated into two blocks, one block with overtly produced primes, and one with inner-speech primes. The order of these blocks was counterbalanced. Each trial consisted of a shrinking–flashing event pair. Each block began with 8 filler trials during which the experimenter monitored the participant's understanding of the task. After completion of the first block (a total of 80 trials), participants switched to the other prime type and continued for another 80 trials. Filler trials contained 4 pictures with dissimilar names from a filler set of 48 images. There were a total of 24 critical trials, with 12 critical trials in the first block, and 12 critical trials in the second block. For any given participant there were six critical items in each of the four conditions crossing repetition and prime production form. Critical trials contained 2 unique images (the prime and target) and 2 images that had also appeared in the filler trials. There were a total of 160 trials. Thus, with 24 critical trials (half of which were repeated mentions), the target was a repetition of the prime on 7.5% of all trials.

The particular trials experienced by each participant were determined as follows. Because there were four potential prime contexts (inner/overt speech × repetition), four lists of trials were created to counterbalance the items across subjects. To minimize order effects, each list was also reversed, yielding eight lists. In Kahn and Arnold (2015), participants were told to name the *object* in their heads silently (''alligator''), while in this study participants were told to produce an entire utterance (e.g. ''The alligator is shrinking'') either as inner speech or overtly depending on the block. The design in the present study makes the prime sentences more analogous to conversational speech.

Procedure

Participants were instructed to describe the events, e.g. ''The alligator shrinks'' or ''The clown flashes'' following a prompt that appeared immediately after the termination of each shrinking or flashing event, which started the recording. During the overt speech block, participants said the prime sentence aloud. In the inner speech block, participants were instructed to say the prime sentence as inner speech such that they could ''hear'' the sentence in their head, but without moving their mouths. Due to the blocked nature of the experiment, it was always clear to the participant what type of production was required for the prime. The target phrase was always said aloud. Once they completed the description of each event, participants clicked the mouse to end the recording and move on to the next event.

Results

The key results concern the target word onsets and durations across conditions. Target word onsets were defined as the length of time from the beginning of the recording to the target word, including the determiners when participants produced them. The duration of the word ''the'' in the target sentences is therefore included in target word onset time. Durations are defined as the entirety of the target word, from the end of ''the'' to the beginning of ''flashes''.

Coders completed target segmentation by hand in Praat (Boersma & Weenink, 2014) using the spectrogram and the waveform. Coders were blind to the experimental conditions that generated the target trials. This was accomplished by presenting each utterance to coders in isolation. All recordings were shuffled and renamed so that coders could not know from what condition the target utterance originated. Coders were told to segment out the target word such that it was not identifiable as anything other than the target, did not contain sounds from the article such as the schwa, and were taught the spectrogram forms for transitions between fricatives ([s] to [f], [ʃ] to [f]). To check the reliability of the coding system, a second coder (the first author) separately coded all of the trials of three participants (a total of 792 trials). These were compared to the primary coding results, taking the absolute value, yielding an average difference of 2 ms (with standard deviation of this measure of only 3 ms) in total duration for the target word.

For statistical hypothesis testing, the target word onsets and durations were log transformed, then centered with respect to the grand average and scaled according to the standard deviation of the entire sample. Due to the log transform, the reported coefficient estimates resulting from the analyses do not directly translate to the reported millisecond differences. The models used for analysis are maximal mixed effects models with random slopes and intercepts by items and by participants. The random slopes in these models are those justified by the 2-by-2 factorial design, which in this case includes repetition, prime form, and their interaction. These factors were sum coded. Models were built using the R package lme4 version 1.1-7 (Bates, Maechler, & Bolker, 2012). We take t values to approximate z scores and assume that absolute values above 1.96 are significant in a two-tailed test. All significance levels represent a two-tailed test unless otherwise specified.

Only target (''flashing'') utterances are considered in these analyses, and only those where the name matched the most common label in the Snodgrass and Vanderwart (1980) materials were included. Whenever the prime or the target ''rabbit'', ''refrigerator'', and ''grasshopper'', was called by an alternate name (e.g. ''bunny'', ''fridge'' and ''cricket'', respectively), the entire trial (both the prime and the target) was rejected. If speakers made a disfluency, restarted, or mispronounced the prime or target, the target was excluded as well. Additionally, when the participant said the prime sentence aloud but should have said it as inner speech, the corresponding targets were excluded from the analyses. Overall, 101 out of 1320 total trials meeting these criteria were removed.

We first consider target word onsets. Repeated targets were produced earlier than non-repeated targets, (668 ms versus 732 ms; $\beta = -0.08$, $t = -3.13$). Target word onsets in the inner speech condition were significantly later than in the overt speech condition (773 ms versus 628 ms; $\beta = -0.16$, $t = -4.30$). There was no interaction between these two factors ($\beta = -0.03$, $t = -0.71$), suggesting that repetition speeds up target word onsets to a similar extent regardless of the production form of the first utterance. These results are reported in Table 1.

To examine the effects of the manipulations on word duration, we again built a model to test for main effects of repetition, prime sentence form, and their interaction. There was a significant interaction between prime form and repetition, such that only targets following overt primes were reduced ($\beta = -0.15$, $t = -2.14$). Repeated targets following overt primes were shorter than repeated targets following covert primes (17 ms shorter versus 6 ms longer). There was no main effect of repetition (435 versus 440 ms; $\beta = -0.09$, $t = -0.50$). These findings are numerically similar to those of Kahn and Arnold (2015), who found a 2 ms repetition reduction effect for duration following inner speech primes, but a 22 ms effect for overt primes. These results are reported in Table 2. Figs. 1 and 2 depict the results for target word onsets and durations in this experiment, respectively.

Discussion

The results from Experiment 1 suggest that repetition in unmouthed inner speech and overt speech leads to reduced target word onset times, but only repetition in overt speech leads to reduced duration. Overall, these data confirm some aspects of the facilitation reduction hypothesis. Priming the production system does have consequences for the speed of utterance production. However, priming different levels of the production system has different consequences for target word onsets and durations. The fact that target words are reduced in duration following repeated overt primes suggests that a non-conceptual, possibly audition- or articulation-mediated priming mechanism within the production system governs word duration. Thus, the lack of reduction for repeating referents after inner speech is inconsistent with a theory in which repetition reduction of a word results from facilitation at any level of the production system.

Table 1
Fixed effects for Experiment 1 onsets.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.16	0.04	−4.30
Repetition	−0.08	0.03	−3.13
Prime form × Repetition	−0.03	0.04	−0.71
(Intercept)	0.03	0.14	0.21

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

Table 2
Fixed effects for Experiment 1 durations.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.14	0.08	−1.84
Repetition	−0.09	0.18	−0.50
Prime form × Repetition	−0.15	0.07	−2.14
(Intercept)	0.03	0.14	0.21

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

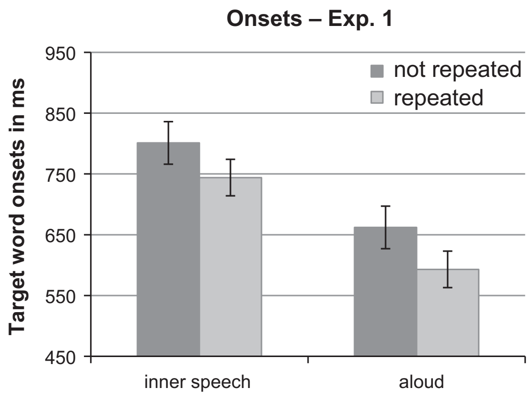

Fig. 1. Experiment 1: Repetition reduces target word onsets. Error bars represent one standard error as assessed by the statistical model.

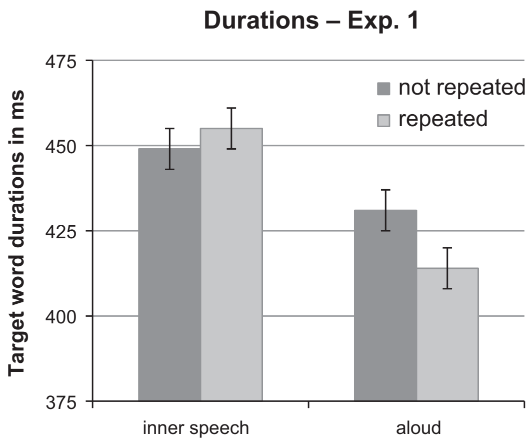

Fig. 2. Experiment 1: Repetition shortens durations after overt speech. Error bars represent one standard error as assessed by the statistical model.

One potential explanation for the lack of an effect of repetition on duration in the inner-speech conditions is that participants did not actually retrieve the lexical labels of the primes. That is, they did not produce the primes inwardly as instructed. However, we think this is unlikely because the repetition priming effect for target word onsets in the inner speech condition was as robust as the effect in the overt speech condition. Studies of repetition priming effects in naming have concluded that object recognition and conceptual identification processes contribute additively with processes associated with lexical access, and moreover that lexical repetition contributes the lion's share of whole repetition priming effect (Francis, Corral, Jones, & Sáenz, 2008). Given this, the near identity of the repetition priming effects for target word onset time in the inner and overt conditions must involve the contribution of priming from lexical access. Thus, we can rule out the possibility that differences in repetition priming for duration are the result of participant non-compliance.

An additional concern is that the verbs might be accented in such a way that would interact with the durations of target words. To assess this possibility we ran an additional model looking at the prosodic features of the word "flashes" as a function of repetition, prime form, and the interaction of these. We examined mean F0, max F0, and verb duration. Both verbs said after repeated inner speech and repeated overt speech were equally accented (150 Hz) and similar in duration (779 versus 766 ms, respectively). We found no statistically significant differences across the groups for any of the prosodic measures on the verb, suggesting that the verbs in the target clause were not differentially accented due to the experimental condition they were in. Given all of these considerations, the most straightforward account of the results is that the overt articulation of the noun or its auditory consequence is necessary for reduction when it is repeated.

In Experiment 2, we examined whether or not having access to the articulators and the motor plans for articulatory gestures is critical. Participants performed a mouthed inner speech variant of Experiment 1. Mouthed inner speech shares more of the articulatory phonetic properties of overt speech, but lacks sound (Oppenheim & Dell, 2010). If such representations play a role in repetition reduction, we should find repetition reduction effects for both target word onsets and durations when the primes are mouthed.

Experiment 2

Method

Participants

Participants were 30 native English speakers from the University of Illinois Paid Subject Pool. Each person received $8 for their participation in the study.

Procedure

The materials and design were the same as in Experiment 1. However, there was a difference in the task. Participants in the inner speech stages of the experiment were told to mouth the prime sentences, as if they were trying to tell someone what had happened from across the room without letting others know.

Results

The models were built in the same way as Experiment 1. We removed trials where participants did not name the object with its prototypical label or produced the prime aloud in the inner speech condition, resulting in the loss of 105 trials.

As in Experiment 1, we first examined the effect of repetition on target word onsets. Repeated targets had earlier target word onsets than non-repeated targets (620 ms versus 680 ms, $\beta = -0.05$, $t = -2.05$). Target word onsets following mouthed inner speech primes were later than targets following overt primes (685 versus 615 ms), though this difference was only marginal ($\beta = -0.10$, $t = -1.81$, $p < .10$). The target word onset results are shown in Fig. 3 and summarized in Table 3. As in Experiment 1, there was no interaction ($\beta = 0.02$, $t = 0.30$).

As in Experiment 1, repetition reduction was greater in the condition in which the primes had been said in overt speech (15 ms shorter in duration for targets following overt versus 0 ms shorter for targets following covert primes), but unlike Experiment 1, the relevant interaction was not significant. Given the findings of Experiment 1, we tested for the predicted reduction in duration in the overt condition as a result of repetition, and found a marginal effect in the expected direction ($\beta = -0.035$, $t = -1.58$, $p = .06$, one-tailed test). These results are shown in Fig. 4 and summarized in Table 4.

Experiments 1 and 2 cross-experiment analysis

Although the effects of Experiments 1 and 2 were quite similar in pattern, the effects of repetition on duration in Experiment 2 were not statistically robust. In order to assess the robustness of the effects and compare the differences in the effects across the experiments, we analyzed both experiments together with experiment as an additional fixed-effect interaction term. We assessed the role of repetition, prime form, and the interaction between these two factors across the two experiments. For target word onsets, the main effect of repetition was preserved, ($\beta = -0.13$, $t = -2.77$), with earlier target word onsets in utterances containing repeated targets. Similarly, we found that the main effect of prime form was still robust, such that target utterances following covert primes led to significantly later target word onsets ($\beta = -0.13$, $t = -5.06$). There was no interaction between these factors. We considered the possibility that the effects of prime form and repetition might vary by experiment. The experiment factor did not interact with any of the fixed effects or their

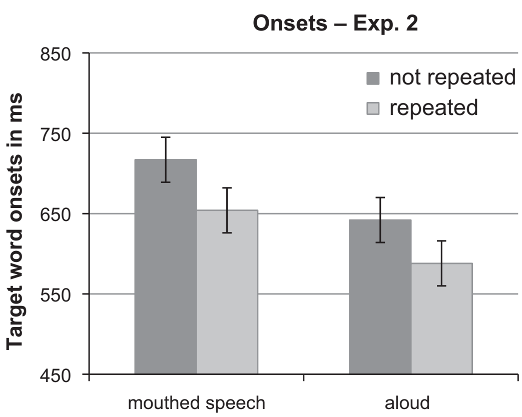

Fig. 3. Experiment 2: Repetition reduces target word onsets. Error bars represent one standard error as assessed by the statistical model.

Table 3
Fixed effects for Experiment 2 onsets.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.10	0.05	−1.81
Repetition	−0.05	0.03	−2.05
Prime form × Repetition	0.02	0.05	0.30
(Intercept)	0.01	0.13	0.09

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

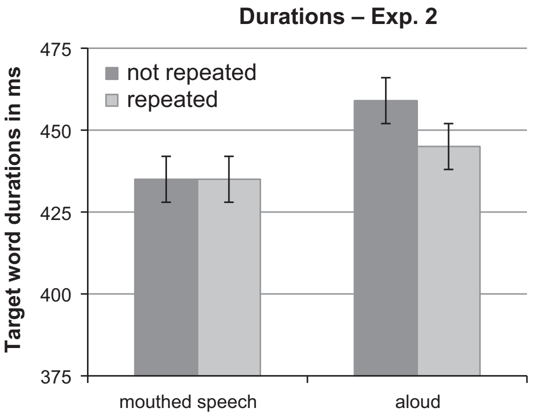

Fig. 4. Experiment 2: Repetition shortens durations after overt speech. Error bars represent one standard error as assessed by the statistical model.

Table 4
Fixed effects for Experiment 2 durations.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.03	0.08	−0.33
Repetition	−0.04	0.04	−0.97
Prime form × Repetition	−0.05	0.07	−0.72
(Intercept)	0.01	0.13	0.09

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

interaction. The model for target word onsets is reported in Table 5.

The analogous model for durations demonstrated that durations were in general longer following inner speech primes, an effect largely due to the unmouthed inner speech in Experiment 1 ($\beta = -0.14$, $t = -2.71$). The critical interaction between prime form and repetition was robust, such that effects of repetition on duration reduction were greater for targets following overt primes than inner speech primes ($\beta = -0.08$, $t = -2.65$). There was no main effect of repetition on durations ($\beta = -0.01$, $t = -0.42$). The magnitudes of the fixed effects and their interactions did not vary as a function of experiment. The model for durations is reported in Table 6.

Discussion

In Experiment 2, we tested whether repetition in mouthed speech led to reduction of the repeated words. Although repetition led to earlier target word onsets in both mouthed and overt speech, reduction in target word duration only occurred for repetition in overt speech.

Table 5
Fixed effects for Experiment 1 and 2 onsets.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.13	0.03	−5.06
Repetition	−0.13	0.01	−2.77
Prime form × Repetition	−0.004	0.02	−0.30
Experiment (1 baseline)	−0.04	0.03	−1.57
Experiment × Prime form	0.04	0.03	1.47
Experiment × Repetition	0.01	0.01	1.31
Experiment × Prime form × Repetition	0.01	0.02	0.54
(Intercept)	−0.07	0.14	−0.49

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

Table 6
Fixed effects for Experiment 1 and 2 durations.

Predictor	Estimate	*SE*	*t*
Prime form (overt baseline)	−0.14	0.05	−2.71
Repetition	−0.01	0.03	−0.42
Prime form × Repetition	−0.08	0.03	−2.65
Experiment (1 baseline)	0.02	0.07	0.27
Experiment × Prime form	0.07	0.05	1.50
Experiment × Repetition	0.01	0.01	0.35
Experiment × Prime form × Repetition	−0.01	0.03	−0.34
(Intercept)	0.002	0.11	0.02

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

Together with Experiment 1, these results suggest that, on the whole, repetition of a word in inner speech does not lead to reduction on the target words. Furthermore, they suggest that even priming aspects of the production system that are engaged in articulation is not sufficient for repetition reduction. The data from these experiments suggest that the word must actually be uttered in order for repetition reduction to occur. In both Experiments 1 and 2, uttering the word led to reduction, while priming the production system up to the point of articulation did not. Why might this be?

If we put these results together with the findings by Kahn and Arnold (2012, 2015) and Bard et al. (2000) that hearing a word spoken by another person can lead a speaker to reduce, we can hypothesize that the critical feature for reduction is the existence of a recent auditory record of the word. Such an auditory record may stimulate the speaker to infer that any potential audience has recently heard the word and hence it can be reduced without penalty to comprehension (Galati & Brennan, 2010).

Alternately, the auditory record may play a more automatic role in the production process. Perhaps if a word's form has been recently heard for whatever reason, it causes reduction (e.g. Kahn & Arnold, 2015). Others have proposed that memory that a word has been said permits speakers to produce less intelligible word forms (Bard et al., 2000). We know that, in general, there are strong interactions between speech perception and production (e.g. Tian & Poeppel, 2013), which is consistent with the general view that prior audition can affect production. Tian and Poeppel (2013), though, also found that individuals' brains show different patterns of activation when they heard their own productions played back to them than when another person's productions were played twice. This specific finding might motivate an expectation that repetition reduction effects might differ depending on whether the prior auditory experience of the word was from oneself or from another. In any case, we must consider the mechanisms whereby the auditory record affects production.

Later, in the 'General discussion', we will offer a specific hypothesis about how recent hearing of a word could lead to reduction. But for now, it is important to test a prediction of the hypothesis that an auditory record of the word form causes repetition reduction, because this appears to be the simplest possible conclusion from our two experiments, as well as the prior research of Kahn and Arnold (2012, 2015). Consequently, in Experiment 3, we ask whether priming targets with their homophones leads to reduction.

There is some indication in the literature that processing a word later facilitates its homophone mate (Wheeldon & Monsell, 1992), and some that it does not (Fowler, 1988). Furthermore, there is evidence that members of a homophone pair are not necessarily produced identically (Gahl, 2008). Thus there is some uncertainty about whether or not a homophone prime will lead to reduction of its mate. But, if it does, such a result would provide strong support for a repetition reduction mechanism that depends on recent auditory experience with a word's form, rather than the word itself.

Experiment 3

Method

Participants

Forty-eight undergraduate students from the University of Illinois participated in exchange for course credit. All of the participants were native English speakers.

Materials and design

A total of 142 images were used. A subset of the images (70) was taken from the colorized versions of the Snodgrass and Vanderwart (1980) drawings by Rossion and Pourtois (2004) and the rest were Clipart images. The stimuli consisted of 21 homophonous pairs (e.g., *pie-pi*), which were the targets in the critical trials. Fifty additional pairs were assigned as filler stimuli that either partially overlapped in phonology (e.g., *beaker-speaker*, *candy-candle*) or were unrelated phonologically (e.g., *apple-bed*).

As in Experiments 1 and 2, images for each trial were presented in a 2 × 2 display. On every trial, one of the objects would shrink (the prime), and then one of the objects would flash (the target). All images were presented twice in the experiment, once as a shrinking or flashing object, and once as an unrelated distractor item in the display. The locations of the shrinking and flashing objects were randomized. An example trial layout is available in Fig. 5.

There were 71 trials in total. The 21 critical trials were divided into three conditions, with 7 trials in each. On *homophone* trials, the target object was preceded by its homophone prime. On *given* trials, the target object was a repeat of its prime, and on *new* trials, the target did not overlap phonologically with its prime. Examples of utterances from each of the conditions are given below:

(a) Homophone: The pie shrinks. The pi flashes.
(b) Given: The pi shrinks. The pi flashes.
(c) New: The house shrinks. The pi flashes.

To reduce the salience of the homophone pairs, the targets in the 50 filler trials were preceded by primes that partially overlapped in phonology (13), were repeats (13), or were phonologically new (14). This results in approximately 30% of all trials containing a homophone mention of any type. Note that on trials that do contain the mention of a homophone as the target, the odds of either homophone in the pair being the target were equal independent of the preceding prime. For example, *pie* and *pi* both have a 50% chance of being the referent in the target utterance regardless of whether the prime was a repeat, a homophone, or a new word. Thus, the predictability of the target

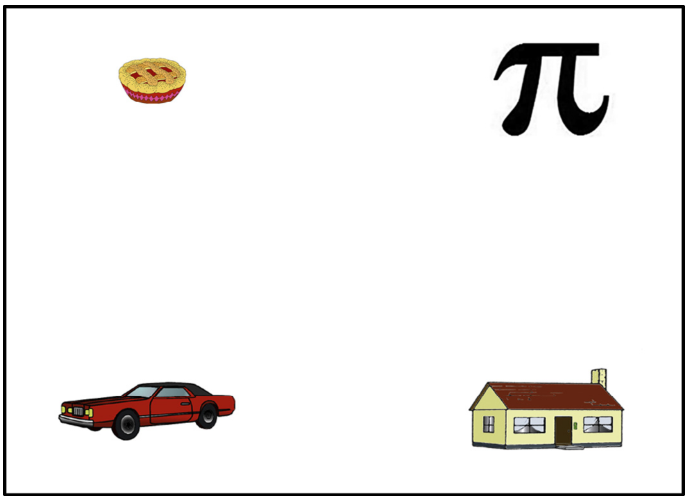

Fig. 5. Schematic of screen for Experiment 3 scenes showing a shrinking *pie* before the production of *pi*.

word form is exactly the same for the three conditions above. The order of the trials was pseudo-randomly permuted, such that two critical trials never appeared in succession. The critical items were counterbalanced using a Latin square, resulting in three lists. The homophone prime–target pairs were also reversed to create three more lists, resulting in six lists in total.

Procedure

To maximize the correct naming of the targets, participants were familiarized with the stimuli prior to the experiment. During this training phase, participants viewed each image individually with its name printed to the side. Participants were encouraged to use the names they had seen in the training phase to describe the shrinking and flashing objects in the experiment.

The experiment began with three practice trials, of which two of the targets were repeated mentions and one was a non-repeated mention. The images in the practice trials were not used in the experimental trials. At the beginning of a trial, four images appeared on the screen. After 2 s, the prime image would shrink. Then, after 1 s, the target image would flash. Unlike in Experiments 1 and 2, in which participants did not begin their utterances until prompted, participants could describe each event as soon as they knew what was happening. All trials were recorded from the start of the scene.

Results

Due to the differences in the methods between Experiments 1 and 2 and this experiment, we can only present analyses for target word durations. Approximately 7.5% of the trials were excluded because the target was misnamed or produced with a disfluency. As in the previous analyses, we first built a maximal mixed effects model with scaled log target word durations as the outcome variable.

Target word durations that were pure repetitions of their primes (e.g. *pi-pi*) were on average 384 ms long. Target word durations following homophone primes (e.g. *pie-pi*) were 429 ms, and those following completely new words (e.g. *house-pi*) were 452 ms long. These results are summarized in Fig. 6. We used dummy coding to compare the homophones to the given and new conditions. We found that when the target words were repeated, they were significantly shorter than when they followed their homophone primes ($\beta = -.33$, $t = -6.04$). However, target words were nevertheless significantly shorter when preceded by a homophone than a phonologically unrelated prime ($\beta = .13$, $t = 2.51$). The model for this analysis is reported in Table 7.

We also examined the pitch of the targets in order to see whether homophones were pitch accented relative to the given condition; we expected this because the homophones are lexically and conceptually new. Pitch accents are correlated with lengthening and hyper-articulation (see Wagner & Watson, 2010 for a review), and we also know that multiple factors can influence a word's prosody. Thus, it is possible that homophone targets were longer than one would expect from the fact that its form was

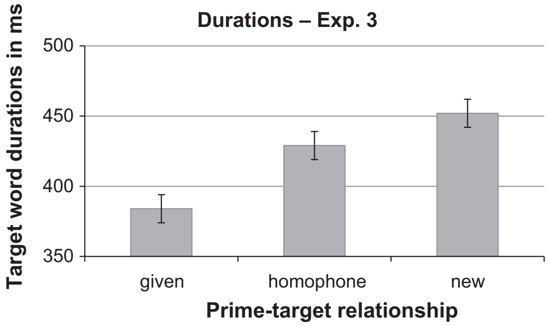

Fig. 6. Experiment 3: Repetition shortens durations after phonological experience. Error bars represent one standard error as assessed by the statistical model.

Table 7
Fixed effects for Experiment 3 durations.

Predictor	Estimate	*SE*	*t*
Lexical repeat versus homophone	−0.33	0.06	−6.04
New versus homophone	0.14	0.05	2.51
(Intercept)	0.08	0.12	0.67

t values over 1.96 are significant to $p < .05$ with a two-tailed test.

primed because it was new and received pitch accenting, reflecting potentially two processes that lead to reduction of repeated words. In sum, perhaps the homophones are shorter than new words because of form priming, and longer than given words because of pitch accenting. Examining pitch provides evidence that homophones receive accenting not evident in duration differences. Both the new targets and the homophone targets differed from the given targets in F0. Homophones had a significantly higher maximum F0 than given words (287 Hz versus 269 Hz; $\beta = -17.84$, $t = -1.72$, $p < .05$, one-tailed test), but were just as accented as new words (277 Hz; $\beta = -10.5$, $t = -0.95$). Homophone and new words' mean F0 was marginally higher (181 Hz and 180 Hz, respectively) than the given target (176 Hz; $\beta = -3.98$, $t = -1.59$, $p = .06$, one-tailed), but homophones and new words were not different from each other ($\beta = -1.05$, $t = -0.42$). The effects are not large, but they suggest that both new words and homophones were more pitch accented than given words. Thus, the duration of the homophone target may be increased by accenting and decreased by priming of its auditory form. The new condition, in contrast, has only the accenting increase, and the given condition has only the form priming decrease.

Discussion

The results of Experiment 3, in particular the 23 ms reduction for homophone primes, support the prediction from the view that recent auditory experience of the word form leads to reduction in duration. They suggest that it is not necessary to have lexical or conceptual overlap to achieve reduction. This result is consistent with other demonstrations of influences of homophone primes on subsequent homophones (e.g. Wheeldon & Monsell, 1992).

Our results are in conflict with Fowler's (1988) study that did not find repetition reduction between first and second mentions of homophones. That study used both list reading where participants are prone to pacing their productions, thereby reducing lexical and production facilitation effects on duration (Kello & Plaut, 2000, 2003), and paragraph-length narratives where the homophones were spaced 10–33 words apart (mean = 21 words). In our Experiment 3, however, homophones were always two words apart (the prime verb "shrinks" and the article before the target noun) and speakers spoke spontaneously. As we theorize in the 'General discussion, we believe that having an auditory memory of a phonological sequence drives repetition reduction; thus, the reading task and large, variable amount of intervening linguistic material between homophones very well could have contributed to the lack of reduction found in Fowler's study.

An additional concern is the possibility that duration reduction on the homophone occurred because attention was attracted to the target picture after production of the prime (e.g. Huettig & Hartsuiker, 2010). The homophone target would then be identified earlier than new words, possibly all the way to lexical access, leading to facilitated production and therefore reduction in duration. However, the expectation that this benefit would specifically reduce duration is inconsistent with the earlier experiments. There was no evidence of repetition reduction in the inner speech conditions of Experiments 1 or 2, even though these conditions led to the target being more quickly identified and produced, as seen in the target word onset data. In fact, the results of Experiment 1 and 2 suggest that facilitated conceptual or lexical access does not lead to reduction on the target noun itself. Thus, it is unlikely that incidental attention to the homophone led to the reduction that we see in Experiment 3.

Importantly, the amount of reduction for homophone primes was smaller than the reduction found for lexical repetition, that is, *pi-pi* (68 ms) as opposed to *pie-pi* (23 ms). This suggests that a purely auditory account of reduction will not suffice and hence one must also hypothesize an additional influence, as we did above when we proposed that the homophone condition is also accented and that accent may influence duration. That result aside, the existence of some duration reduction with homophones is consistent with predictions made from Experiments 1 and 2, as well as those of similar studies that have found auditory and articulatory contributions to repetition reduction (e.g. Bard et al., 2000; Kahn & Arnold, 2015).

General discussion

In three experiments, we tested the hypothesis that facilitated retrieval of a lexical item leads to repetition reduction (Gahl et al., 2012; Kahn & Arnold, 2015) by manipulating lexical–conceptual repetition, with and without sound-producing articulation (Experiment 1 and 2) and phonological repetition (Experiment 3). According to the most general formulation of this theory, facilitation at any level of production can lead to reduction on any part of the utterance. We, however, were interested in the factors that lead to repetition reduction on a target noun that was repeated or not (e.g. Kahn & Arnold, 2015; Lam & Watson, 2010, 2014). Experiments 1 and 2 found that the repetition of a word that was previously produced as mouthed or unmouthed inner speech resulted in a reduction in target word onset time, but not target word duration. Only repeated words spoken aloud led to repetition reduction in target word durations. Based on the present results and those of Kahn and Arnold (2015), we specifically predicted that the repetition of phonological forms, rather than lexical items or concepts, should lead to duration reduction on the repeated words. The reduction of a word following the utterance of its homophone in Experiment 3 suggests that repetition reduction can result from prior production of a word's form, even when it constitutes a different lexical item. We acknowledge that, at present, the field lacks a complete understanding of how homophones interact in production. That said, a prediction from our account was that the *duration* of the homophone target would be reduced, and this result was obtained. This experiment, together with Experiments 1 and 2, is consistent with the conclusion that the speaker must have recently heard the word's form for it to be subsequently reduced.

One potential explanation for these results is that speakers modulate the forms of their utterances for the listener (Bard & Aylett, 2004; Galati & Brennan, 2010). Under such an audience design account, the key variable is whether a word is given or new with respect to the discourse as experienced by the listener. Listeners may prefer for given words to be reduced and for new words to be lengthened because doing so evens out the rate at which information is conveyed (Aylett & Turk, 2004). Alternatively, because the word is active in the discourse, speakers can exert less effort and produce a less intelligible, reduced version of a given word (Aylett & Turk, 2004; Bard et al., 2000; Fowler, 1988). Finally, listeners may use reduction as a cue to givenness because that is how speakers, for whatever reason, tend to produce given words. Consequently, speakers work to match these expectations. If any of these accounts are correct, they explain the lack of reduction in the inner speech conditions of Experiments 1 and 2; because the target in the second mention has not been introduced into common ground, it should not be reduced.

Some of the results, however, contradict an audience design account. First of all, participants in our experiments are not actually addressing an audience in the task, and repetition reduction was still observed. Furthermore, the presence of an audience has been shown in several similar studies to not be necessary for repetition reduction (Kahn & Arnold, 2015; Lam & Watson, 2010, 2014). Of course, it is possible that speakers assume that there is a virtual listener for any speech they produce, and then adjust their linguistic choices accordingly. However, the results from Experiment 3 are more difficult to explain under an audience design account. Under such an account, speakers would *not* reduce or might even be expected to lengthen the second mention of a homophone pair to signal a contrast (Baese-Berk & Goldrick, 2009; Buz, Jaeger, & Tanenhaus, 2014).

Perhaps speakers instead engage in what we call "audience design by proxy". Speakers may intend to reduce for their listeners, but they may have trouble keeping track of the history of the conversation as well as the perspective of the listener (Epley, Keysar, Van Boven, & Gilovich, 2004; Keysar & Henly, 2002), or they may even not remember whether they or their partner first said the word (Bard et al., 2000). In contrast to audience design accounts, Bard et al. (2000) proposed that repetition reduction in duration is due, not to audience design, but rather speaker-internal processes. Speakers who have produced something once and believe it to have been understood can exert less effort during the second production.

Speakers may use their own auditory memory for whether a word was recently produced as a proxy for direct knowledge that their conversation partner heard a word (see a similar proposal at the discourse level by Keysar & Henly, 2002). This view explains the inner speech findings for Experiments 1 and 2. In the inner speech conditions, the speaker has no auditory memory of the prime word. If an auditory memory functions as the speaker's model of the listener's representation of the discourse, they would fail to reduce targets in both mouthed and unmouthed inner speech conditions. The results from Experiment 3 also support this view. In that study, homophones are reduced because the speaker has an auditory memory of saying *pie*, but their memory is also consistent with saying *pi*. Because the speaker consults their own auditory memory for whether a word was introduced into the common ground or not, the speaker fails to keep the homophone word durations as long as if they had been a first mention. In the final section, we introduce the possibility that the online processing of auditory feedback enables speakers to access their (potentially faulty) memory of the entities that have been introduced into the discourse, and to have that access affect word durations.

Facilitation of production by priming the online processing of auditory feedback

Auditory feedback plays an important role in controlling speech production (e.g. see Guenther, 2014; Hickok, 2014, for recent reviews). Not only is feedback used to learn how to correctly produce sounds, it is also used in an immediate, on-line fashion to keep track of where the production process is (update the "internal model") and detect and correct for deviations between expected and actual feedback. The effect of delaying auditory feedback leading to lengthening and disfluency (e.g. Fairbanks, 1955) is a classic demonstration of feedback's role in production.

We propose that repetition reduction is controlled in part by auditory feedback. When the feedback is processed as expected, the production process moves rapidly through its planned sequence of states and the word is produced with a correspondingly short duration (see Fig. 7, which represents these states as phonemes, but that is not a necessary assumption). When the feedback is less efficiently processed, the production process will move more slowly and the word will be longer.

We further propose that a recent auditory memory for the form of the word can increase the efficiency of feedback processing, thus leading to reduction. The auditory memory can come after the overt production of a word, or from having overheard the word from someone else. This mechanism can explain why our first two experiments

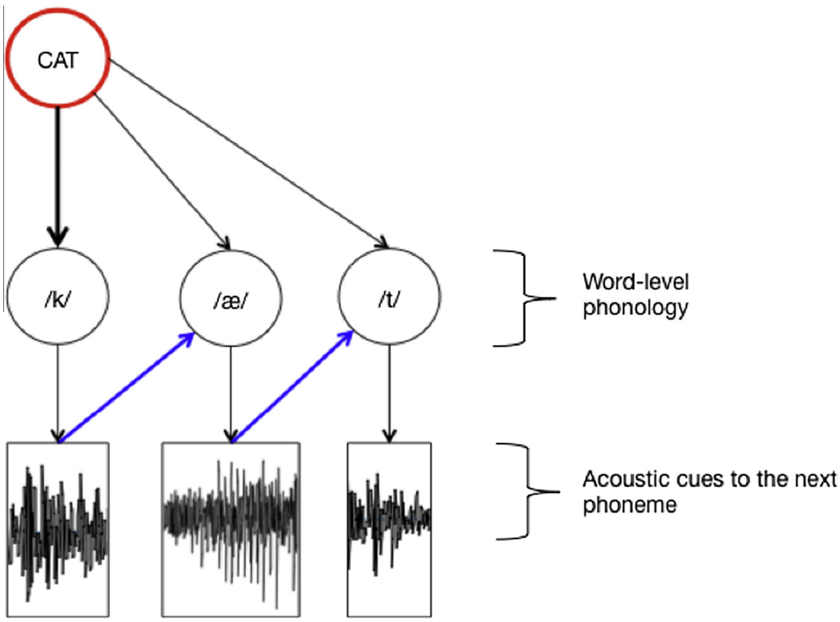

Fig. 7. Schematic of sequential ordering in word repetition. In this model, each stage in the phonological output depends on the phonetic output. In general, each phoneme waits for enough evidence that the previous phoneme has been produced. The next phoneme in a sequence is determined by the word's phonological structure. After a phoneme has been produced, the links between word-level phonology and the acoustic targets cue the beginning of the next phonological segment. When a word is repeated, the activation of the concept is stronger, as well as the connections between the acoustic output and the next phoneme of the word.

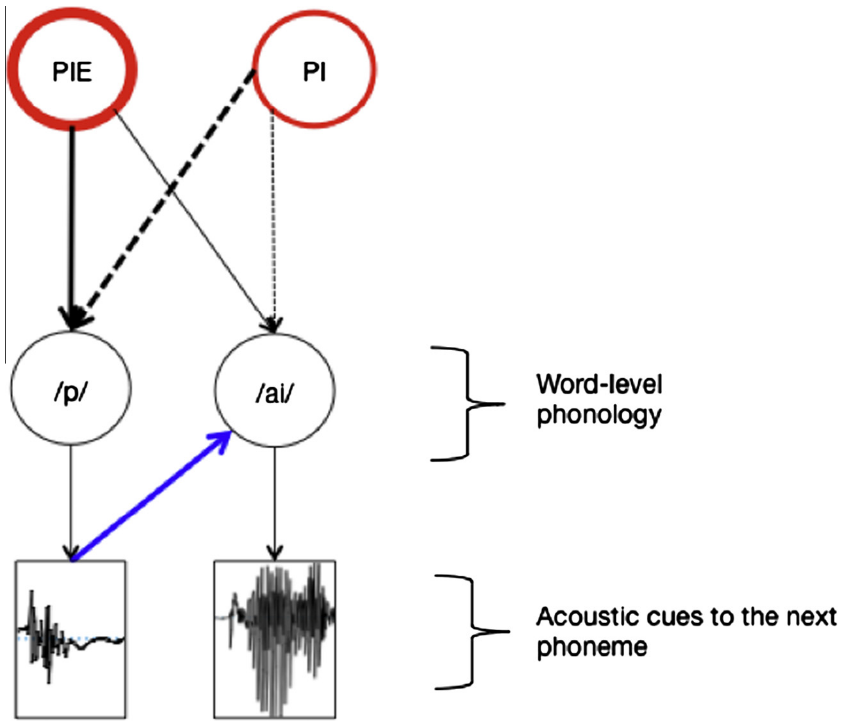

Fig. 8. In the case of homophones (e.g. *pie* and *pi*), we assume identical phonological representations (/pai/). When *pie* is produced, the transitions between the same phonemes relevant for *pi* are strengthened. However, because *pi* is a different word, conceptual or semantic priming cannot occur, leaving it only with the strengthened acoustic cues from the first word, *pie*.

found that repetition reduced target word onsets, but not target word durations, in the inner speech conditions. The prior production of a word by a speaker, regardless of whether it was aloud, leads to speeded access to the beginning of the word (i.e., getting to the starting state in the production of *cat* in Fig. 7). This prior production can be assumed to prime the pathways that lead to the starting state and hence faster target word onsets.

The duration of the word, however, is controlled less from the efficiency of these higher pathways, and instead by the efficiency of the processing of the auditory feedback. When the word has recently been heard, the feedback-guided transition process between the phonological-acoustic states is rapid, leading to shorter durations. When there is no auditory memory of the prior production, as in the inner speech conditions of our studies, there is consequently no reduction from the prior production. That is, this process is not primed. An auditory memory of *pie*, which shares some phonological representations with its homophone, speeds up the processing of feedback for *pi*, leading to partial reduction on the homophone (see Fig. 8), which also receives accenting by virtue of being new to the discourse.

To conclude, the data support the facilitation reduction hypothesis, but in a very specific way. Reduction of repeated words is caused by facilitation (priming) of the production component that uses sound to guide articulation. The factors that influence higher planning levels, such as lexical access or conceptual familiarity, come into play prior to the articulation of the repeated words themselves, contributing to the shortened target word onsets (including the duration of the pre-target determiner; Arnold, Hudson Kam, & Tanenhaus, 2007). We supported these claims in two ways, first by showing that silent production, even when it is articulated (mouthed), does not lead to reduction on the target word, and second, by showing that repetition of a homophone leads to reduction, as we predicted. These effects arise, we propose, from the natural processing of the production system, possibly through the processing of auditory feedback. We further speculate that these processes create a kind of audience design by proxy. Speakers do not have to explicitly track what their audience has heard in order to reduce appropriately. Rather, their own auditory memory does the trick and, if our feedback explanation is correct, this memory automatically creates the reduction.

Acknowledgments

The first author is supported by an NSF Graduate Research Fellowship. This project was funded in part by NIH DC-000191 and R01 DC008774 as well as the James S. McDonnell Foundation. We would like to thank our research assistants Jace McNett, Kathy Kastanes, Steven Anderson, and Xiaomeng Li for their help collecting and processing the data, as well as Matt Goldrick for his comments on this manuscript.

Appendix A

Table 1		Table 2	
Items – Experiment 1 and 2		Items – Experiment 3 Homophone A	Homophone B
alligator	grasshopper	arc	ark
apple	hammer	base	bass
axe	harp	capital	capitol
ball	house	colonel	kernel
barn	key	cue	queue
bat	lettuce	currant	current
bed	motorcycle	cymbal	symbol
bell	peach	doe	dough
brush	pear	fairy	ferry
button	pencil	flour	flower
candle	piano	genes	jeans
car	pipe	hair	hare
chicken	pot	knight	night
church	rabbit	leak	leek
clock	refrigerator	moose	mousse
cow	rooster	muscle	mussel
doll	ruler	oar	ore
eagle	shoe	pi	pie
fence	star	plain	plane
flag	tie	rain	rein
football	tomato	stake	steak
fox	truck		
frog	turtle		
glove	watermelon		

References

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. K. (2007). If you say thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914–930.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*, 527–554.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Bard, E. G., & Aylett, M. (2004). Referential form, duration, and modeling the listener in spoken dialogue. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173–191). Cambridge: MIT Press.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 83–102.

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer [Computer program]. Version 5.4.04*. <http://www.praat.org/> Retrieved 28.12.14.

Buz, E., Jaeger, T. F., & Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyper-articulation. In *Proceedings of the 36th annual meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders, 20*, 333–346.

Fowler, C. A. (1988). Differential shortening of repeated context words produced in various communicative contexts. *Language and Speech, 31*, 307–319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of repetition priming components in picture naming. *Journal of Experimental Psychology: General, 137*, 566.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474–496.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789–806.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.

Guenther, F. H. (2014). Auditory feedback control is involved at even sub-phonemic levels of speech production. *Language, Cognition, and Neuroscience, 29*, 44–45.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*, 2–20.

Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes, 25*, 347–374.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67*, 311–325.

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language and Cognitive Processes, 30*, 103–119.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 719–750.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48*, 207–232.

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13*, 207–212.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137–1146.

Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 829–843.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528–537.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition, 38*, 1147–1160.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6*, 174–215.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience, 25*, 1020–1036.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*, 905–945.

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology, 44*, 723–761.