Speaker-internal processes drive durational reduction

Jason Kahn & Jennifer E. Arnold

University of North Carolina at Chapel Hill

Address correspondence to:

Jason Kahn

CB#3270 Davie Hall

Chapel Hill, NC 27599-3270

Tel: (919) 962-5082

Fax: (919) 962-2537

jmkahn@email.unc.edu

Author Note

This research was supported by: NSF grant BCS-0745627 to J. Arnold. We gratefully acknowledge the assistance of Molly Bergeson, Andrés Buxó, Kellen Carpenter, Kayla Finch, Sam Handel, Isaac Hellemn, Jennifer Little, Leighanne Mcgill, Michael Parrish, Giulia Pancani, Alyssa Ventimiglia, Liz Wagner, in both running the experiments and transcribing the data.

Abstract

Speakers tend to reduce the duration of words that they have heard or spoken recently. We propose that reduction in word duration occurs because of facilitated language production processes. This contrasts with an account of reduction based on speakers' adjustment to their listener's mental state. In three experiments, we ask speakers to give instructions to listeners about how to move objects. In Experiment 1, the instruction was preceded by an auditory prime, which elicited reduced spoken word duration, irrespective of what the speaker thought the listener knew or heard. In Experiments 2a and 2b, we tested the related hypothesis that the speaker's experience articulating the target leads to facilitation above and beyond activating the target word through other means. While prior articulation of the target led to greater reduction than just thinking about the word (Exp. 2a), saying the prime led to equal reduction as hearing the prime (Exp. 2b). Although the manipulated prime only involved the target objects, reduction also appeared on planning regions of the utterances, lending further support to an account based on facilitated processing. The pattern of results leads to the conclusion that facilitated processing, perhaps brought on primarily as a function of having heard a word, explains durational reduction in running speech.

Keywords: Reduction, Audience Design, Facilitation, Speaker-internal processing, Word Duration

Speaker-internal processes drive durational reduction

People often repeat words in conversation, but the pronunciations of these words tends to differ. For example, a teacher might ask his students, "What do the introduction, the body, and the conclusion of a paper do?" And they might reply (in unison!), "The introduction lays out the problem, the body provides evidence, and the conclusion gives the take-home message." Each repeated word is likely to be shorter in duration, lower in pitch or intensity, or some combination of these, even after accounting for baseline differences between the two speakers. This is called acoustic reduction, and our goal in this paper is to examine why it happens.

A well-established generalization is that acoustic prominence (e.g. accenting or reduction) correlates with the information status of the referent. (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon & Newlands, 2000; Fowler & Housum, 1987; Halliday, 1967). Information that is already evoked (usually by having been previously mentioned) is termed **given**, and typically is referred to with shorter and less acoustically prominent pronunciations than **new** (unevoked) information. Even within the given category, acoustically reduced forms tend to be used for referents that are relatively more accessible or salient in context (Ariel, 1990; Gundel, Hedberg & Zacharski, 1993).

However, there is less consensus about why acoustic prominence variation patterns with information status, nor is there any agreed-upon psychological mechanism. Here we examine explanations from two theoretical approaches: 1) the audience design approach, and 2) the processing facilitation approach.

The audience design approach suggests that reduced forms result from the speaker's recognition that the referent is given/accessible in common ground – i.e., in the representation of the discourse that the speaker shares with the addressee (Clark & Marshall, 1981). Because of

this, she can count on the listener needing less explicit information to establish or maintain reference effectively (Chafe, 1976; Clark and Haviland, 1977; Lindblom, 1990). Under this view, the speaker reduces subsequent mentions to an entity *because* of his representation of the listener, under the assumption that the shared knowledge reduces the need for explicit information in the utterance. A related view suggests that information status serves as the input to pragmatic rules about when it is appropriate to use reduced forms (Grosz & Sidner, 1986; Hirschberg & Pierrehumberg, 1986; see Venditti & Pierrehumbert, 2003, for a review). Although these models do not necessarily involve an explicit representation of the listener's perspective, information status is typically defined in terms of common ground (e.g. Gundel et al., 1993). Under both of these views, the speaker must access a model of the listener's knowledge for acoustic reduction decisions.

An alternate view focuses instead on how the speaker's own experience might lead to the production of acoustically reduced forms for given or salient information statuses. This view builds on evidence that speakers reduce because of facilitation of their language production processes, sometimes without incorporating information about the listener (Aylett & Turk, 2004; Bard et al., 2000; Bard & Aylett, 2004; Bell, Brenier, Gregory, Girand & Jurafsky, 2009; Bybee, 1999; Gahl, 2008; Horton & Gerrig, 2005; Lam & Watson, 2010).

There is empirical support for both audience design and facilitation views, which we outline below. Yet there are also numerous open questions about the extent to which each contributes to variation in acoustic prominence, and in particular about the mechanism. For example, do speakers require a well-established common ground representation to guide reduction, or do they sometimes reduce simply because of internal priming? What exactly does a

representation of common ground contribute to the processes involved in reduction? Does articulation of relevant information lead to reduction, even independently of common ground?

The experiments below attempt to shed light on these questions by focusing on the diverging predictions of audience design and facilitation views about reduction. We specifically test two major predictions of this view, namely that if speakers reduce because of facilitation of production mechanisms, they should do so 1) without regard for their representation of the listener; 2) more after articulating the word themselves than after merely hearing it.

Evidence for audience design effects on acoustic reduction

Audience design refers to the process of modifying an utterance for the sake of one's addressee. In the example above, the students might acoustically reduce the repeated words, because of the information being shared in the conversation. This shared status allows interlocutors to create rich representations of the knowledge they share with listeners, called *common ground* (Clark & Marshall, 1981). Both speakers and listeners appear to use these representations to guide their language use, such as when they choose lexical expressions. For example, conversational participants tend to settle on a single label for an object (e.g., *the pennyloafer* vs. *the shoe*), and persist in using it even if it becomes less appropriate for the discourse situation (Brennan & Clark, 1996). Switching to a new partner after successfully entraining on a label reduces the dependence on this label, which highlights the importance of common ground for everyday language use. Similarly, speakers are more likely to use a simple name for referent when they have good evidence that name is shared with their addressee (Gorman, Gegg-Harrison, Marsh & Tanenhaus, 2012; Heller, Grodner & Tanenhaus, 2008). In comprehension, too, listeners use rich information to guide their interpretation of referential

expressions (Brown-Schmidt, 2009, 2012; Hanna, Tanenhaus & Trueswell, 2003; Hanna & Tanenhaus 2004; Gorman, et al., 2012; Schober & Clark, 1989).

Yet the shared nature of common ground presents a potential challenge to researchers seeking empirical support for its effects: any information that is shared is also available to the speaker alone, and thus could be used egocentrically (cf. Horton & Gerrig, 2005). For this reason, many studies have used paradigms that manipulate the speaker and listener's experience independently. In a map-based cooperation task, for example, Bard & Aylett (2004) compared the first mention of a referring expression to the same speaker's second production of the same referring expression to a second addressee. They found that repeated words were shorter and rated as less intelligible under both conditions, regardless of listener knowledge. The authors proposed that their results are explained by a division between slow, deliberate message-level processing (where lexical expressions are chosen), and fast-cycling lower-level processing, where articulatory decisions are made (see also Bard, et al., 2000). However, the findings from the map-task corpus leave two questions unanswered. First, although the authors showed that speakers reduced even for egocentrically known information; they did not ask whether common ground might matter too, potentially leading to an even greater amount of reduction (Galati & Brennan, 2010). Second, the corpus analysis did not control for the effects of linguistic context. Critically, the amount of time since first mention likely differed between same-addressee and new-addressee conditions.

Further evidence for this processing division comes from a study that manipulated listeners' anticipation of the speaker's message (Arnold, Kahn & Pancani, 2012). We asked participants to give instructions to a confederate to place objects on a colored board. The confederate either anticipated the speaker's instruction by picking up the intended object, or

waited for the speaker to finish the instruction. Despite the salient differences between the confederates' behavior, no differences emerged on either the duration of the object word or its rated intelligibility. The duration of the onset to speak latency and the determiner before the object noun (*the*), however, did differ by condition. This suggests that speakers were not designing the acoustic prominence of their utterance for the listener, but neither were they categorically ignoring her. Of particular interest is that the determiner *the* is sometimes considered a planning region (Clark & Wasow, 1998). Speakers often plan speech incrementally, such that the difficulty of producing the noun can be reflected in earlier regions (e.g. Bell et al., 2009). Thus, this work suggested that the behavior of listeners guided the speakers' attention in a way that facilitated language production.

On the other hand, a few studies do provide limited evidence that speakers vary the acoustic prominence on a word based on common ground. Galati & Brennan (2010) asked participants to describe familiar cartoons three times, to two different listeners. In particular they focused on whether the intelligibility of words diminished on the second and third telling, and whether the degree of this effect depended on whether the addressee was the same as the first story, or a new addressee. This structure provided an additional control that was absent from Bard & Aylett (2004), who compared only first and second mentions. Repeated tokens of referential expressions to the same listener did not differ on duration from tokens spoken to another listener, suggesting no effect of audience design. The extracted tokens did, however, differ on the intelligibility ratings provided by naïve listeners, suggesting that speakers do modulate some aspect or aspects of prosody other than duration.

The strongest evidence that speakers use acoustic reduction for listeners comes from another study, using the same paradigm as Arnold et al. (2012), in which speakers gave

instructions to listeners who were either attentive or distracted with a secondary task (Rosa, Finch, Bergeson, & Arnold, under review, this volume). Speakers pronounced the target word with longer duration, e.g. *teapot* in *Put the teapot on the orange circle.* Notably, distracted addressees did not elicit longer durations across the entire utterance, but specifically on the information that was necessary to initiate the action. This shows that acoustic reduction and prominence can be influenced by the addressee's behavior.

In sum, there is mixed evidence about the role of audience design on acoustic prominence. Some studies show that listener knowledge or attention has no effect (Bard & Aylett, 2004) others show it does (Rosa et al., under review), and others show effects limited to intelligibility ratings but not duration (Galati & Brennan, 2010), or only planning regions (Arnold et al., 2012). The experiments we report in this paper examine this question further by asking comparing the predictions of an audience design account with a facilitation account, specifically focusing on questions about the mechanisms that underlie reduction.

Evidence for facilitation effects on acoustic reduction

There is extensive evidence that fluctuations in ease of processing lead to reduction, both in corpus and experimental studies (Balota, Boland and Shields, 1989; Bard et al., 2000; Bell et al., 2009; Lam & Watson, 2010; for a review, see Arnold & Watson, under review, this volume). There is no current consensus on the specific mechanisms that lead to reduction, but numerous possibilities have been proposed. These suggestions are couched within the framework of processing models of language production, which divide the process of producing an utterance into several stages (Levelt, 1989; Bock 1982; Dell, 1986; Levelt, Roelofs & Meyer, 1999; Garrett, 1980). Typically this begins with a message or conceptualization stage, where speakers organize the general form of their utterance, including the entities and actions involved.

Semantic and grammatical selection follow, where speakers choose the particular lexical entries they will use, and place them into the appropriate syntactic slots. Each word's lexical entry then activates its corresponding phonological form, the abstract sound representations that guide the motor codes that actually produce speech sounds.

Based on current assumptions about model composition (i.e. staged, activation-based processing and selection; Dell, 1986, Levelt et al., 1999), information status could affect acoustic reduction through one of several mechanisms: by easing selection or activation, overall speed of processing from beginning to end, overall level of activation, or several other compatible formulations. Bybee (1999), for example, explains frequency effects on word duration in terms of canalization and fluency of processing. Bell et al. (2009) similarly propose that a mechanism coordinates the lexical and phonological levels of production, lengthening words during online processing to promote fluency. For further discussion of possible mechanisms, see Kahn & Arnold, 2012, and Arnold & Watson, under review, this volume.

We have proposed a specific version of the facilitation approach that we call the Facilitation-based Reduction Hypothesis (FRH; Kahn & Arnold, 2012). The FRH suggests that the degree of durational reduction on a word is related to the degree of facilitation within the language production system, either at a particular level, or across multiple levels simultaneously. Evidence for this claim comes from an instruction-giving task (Kahn & Arnold, 2012), in which a speaker and listener saw an array of 8 objects, each on their own computer screen. The speaker described the movement of 3 objects per trial, as they happened, e.g. *The accordion rotates.* Prior to object movement, both participants saw either a linguistic or non-linguistic prime, or no prime (the control condition). The non-linguistic condition, in which the three objects flashed, elicited durational reduction on the object word relative to the control condition. Importantly, the

linguistic condition, in which the participants heard the names of the objects, elicited even more durational reduction than the non-linguistic condition. The most natural explanation of this pattern of results, we argued, was that the linguistic condition facilitated a greater proportion of the processes necessary for language production than the non-linguistic condition, leading to more reduction.

Overview of experiments

In this paper, we take up two open questions about acoustic reduction, both related to how the speaker's experience articulating a word affects later mention of the same word. The level of articulation offers a window onto the mechanisms of acoustic reduction, because it has the potential to relate to both audience design and production facilitation. When a spoken word enters common ground, speakers and listeners both are likely to activate representations at the conceptual, phonological, and lexical levels. The articulatory motor codes, however, are unique in participating in a process that is activated by the speaker alone. Thus, the FRH predicts that articulation experience should lead to additional acoustic reduction. By contrast, the common ground account predicts that either hearing or speaking the word in the absence of an addressee should lead to less reduction than when the information is shared, or possibly no reduction at all.

We address these questions in two steps. The first experiment manipulates common ground knowledge of an auditory prime, to isolate the effects of facilitation and audience design on acoustic reduction. Because language is communicative, speakers and addressees frequently share common ground. In such situations, speakers might reduce for their listeners, or simply because of the ease of their own processing. Experiment 1 follows earlier work in using a referential task with both common and privileged ground, while controlling for the distance between the facilitating prime and the utterance, and introducing slightly more complex, but still

controlled, utterances. This allows us to ask whether durational reduction emerges because of speaker-internal facilitation, independently of common ground.

Second, we ask specifically about the role of articulation in durational reduction. In Experiment 2a, we ask speakers to produce instructions after hearing a prime or saying it silently to themselves, to test whether facilitation at the articulatory level creates reduction in addition to facilitation at the lexical and/or phonological level. In Experiment 2b, we ask a similar question while incorporating a listener. This allows us to ask whether speakers reduce more after saying a prime themselves than after hearing a listener make the same contribution, in effect isolating the contribution of articulation to reduction during common ground processing.

All the experiments required multi-word instructions like *The airplane fades*, which afforded an opportunity to examine whether reduction appears on words other than the target. This allowed us to test the additional prediction that facilitation effects can appear in planning regions, as well as on the manipulated target.

Experiment 1

We examined first whether acoustic reduction is influenced by the speaker's knowledge about whether the listener heard the first mention. We used the same experimental paradigm as Kahn & Arnold (2012), in which speakers gave short instructions for their partner to move objects on screen, e.g. "*The airplane fades.*" The critical manipulation was an auditory prime before the target, which occurred in four conditions: heard by the speaker only, the listener only, both participants, or neither participant.

This experimental paradigm provides several advantages over Bard & Aylett (2004), which it extends conceptually. First, it allows us to more tightly control the discourse context by holding constant the distance between prime and target across the experiment. This distance was

relatively short (3 sentences), which differs from both Bard & Aylett (2004) and Galati & Brennan (2010). We also require our subjects to use the same sentential frame in all cases, which allows us to compare prosodic variation independently of syntactic variation. And finally, the randomized block design allows us to use speakers as their own controls, and at the same time control for order effects.

Method

Participants. A total of 18 undergraduates participated in an instruction-giving experiment as speakers, and an additional 13 participated as listeners, all for course credit. We excluded 3 of these pairs due to technical difficulties, for a total of 15 speakers and 10 listeners. The unpaired 5 speakers participated with a male confederate lab assistant. Post-experiment questionnaires revealed that none of these speakers suspected the confederate, nor did the results below change as a function of the confederates.

Materials and Design. 198 colorized versions of the Snodgrass & Vanderwart (1980) line drawings served as the stimuli (Rossion & Pourtois, 2001). These drawings were normed by native French speakers for imageability, visual complexity, familiarity, as well as name agreement in French, although this latter measure was not used here. 18 of these objects served as **experimental targets**, and were chosen to have two syllables and similar values on each of the normed dimensions, as well as frequency. The remaining objects served as fillers.

The general structure of a trial was as follows. 8 objects would appear on both participants' screens, including one of the experimental targets and two randomly-chosen **non-targets**. After 1 second, a recorded voice would speak the names of the experimental and non-targets, in randomized order, either to the speaker, the listener, both participants, or neither of them. Primes were recorded by the first author in a regular speaking voice, with no distinctive

prosody (i.e. a simple declarative sentence: *the airplane*), and played through computer speakers. This priming section was immediately followed by a beep that alerted the speaker that an object was about to move. One second after the beep, one of the non-target objects would perform an action, such as rotating. The speaker would then instruct the listener to perform that action, for example saying *The airplane rotates*. The listener would then execute this instruction by clicking on the appropriate action button (here, Rotates) and then the appropriate object (here, the airplane). The same sequence of beep, object movement, and instruction execution followed for the second non-target object, then the experimental target. Possible actions included rotating, shrinking, expanding, and fading, all of which the listener could execute similarly on her screen by clicking.

There were four experimental blocks, which corresponded to four ways in which participants received priming information. In the given-to-both condition, speakers and listeners both heard the names of the target objects spoken aloud at the same time. In the speaker-only condition, only the speaker heard the names of the target objects, over headphones. Similarly, the listener-only condition gave the priming information only to the listener, also over headphones. Finally, in the control condition, participants received no priming information, and the objects merely moved after a delay. All trials within a block were of a single type, and blocks were presented in a random order between participants. Each participant pair thus saw each experimental target once in each condition, for a total of four presentations per pair. This might lead to the concern that speakers would begin to reduce words on later trials simply because they had already mentioned them several times. This concern was mitigated by four factors: 1) the randomization of condition meant that block number and condition were manipulated orthogonally; 2) Kahn & Arnold (2012) found no differences between blocks, in a task that put

less distance between repetitions; 3) the items appeared randomly within blocks; 4) the results reported below do not change with the inclusion of block number as a predictor in the models.

In order to highlight the differences between these conditions, participants were given three pieces of information at the beginning of each block: 1) both were told explicitly who would hear the names of the objects; 2) each participant put on a pair of headphones for blocks in which he heard the names; 3) the speaker saw an icon at the top of her screen that indicated whether or not her partner was hearing the primes (Figure 1). We hoped that the abundance of information about the primes would make the disparity in the speaker-only and listener-only conditions readily apparent.

A customized Python script written by the first author controlled this and all other experiments reported here.

(Figure 1 about here)

Procedure. Participants each sat in front of their own computer, approximately 10 feet from one another, but where each could see only his or her own screen. They were instructed that they would each see the same array of 8 objects (Figure 2), and they would each perform different jobs on each trial. The speaker was told that she would see an object perform an action, and would have to issue an instruction to the listener about that movement. The listener was told he would have to click on the appropriate object, then the appropriate action button to execute the instruction. They were told this would happen three times for each trial. Examples of the possible actions (expand, rotate left, rotate right, shrink, and fade) were then shown with random stimuli, and a sentence frame was provided. For example, the speaker might have seen an airplane rotating, and would have been told "You might see this [the airplane rotating], and you

might tell your partner 'The airplane rotates.'" The listener was told he would then click on the airplane, then the Rotate Left button on his screen to make it rotate.

(Figure 2 about here)

Next, the participants were told that the basic trial structure would be modified on some trials, and one or the other or both of them would hear the names of the moving objects spoken out loud, in no particular order, ahead of time. To help them keep track of this, they were told that before each set of trials they would know who would be hearing the names, and that the person who would be hearing the names would put on headphones. The speaker was also told about the icons, specifically that one icon meant her listener would hear the primes in the current block, and that the other meant she would not. Participants then completed four practice trials of each type, in a fixed order. The practice trials began with the control trials to familiarize them with the general task, then four each in the speaker, listener, and both conditions. Then the participant began the experimental blocks.

Analysis. The third instruction from each trial was analyzed using Praat software (Boersma & Weenink 2009). The first and second instructions were excluded because they were not fully counterbalanced across participants. The first author examined the visual waveform and listened to the sound file to identify the onset of each trial (indicated by a beep that co-occurred with each action onset), as well as the onset and offset of the determiner *the,* the noun, and the action phrase. Here we present an analysis of four duration measures: latency to begin speaking (time from the onset of the beep to the onset of speech), the determiner (*the*), the duration of the noun (target word), and the duration of the action word. The latency to speak and the determiner for each instruction both are likely to reflect planning time, where shorter latencies indicate less complex and/or demanding plans. Utterances were excluded if they were disfluent in any way

(e.g. pauses longer than 250ms, false starts, disfluent 'the') or if the speaker used the wrong object or action word.

For each of the duration measures, we built a multi-level model to assess whether the priming information in each trial type led to different degrees of reduction. The logarithm of each duration was taken to increase its linear relationship with the predictors. Subject-wise outlier exclusion was not necessary after this transformation. Models were constructed with the lmer function from R's lme4 package, which uses Maximum Likelihood estimation to obtain parameter estimates.

Significance is assessed under the assumption that a sufficiently large dataset (n = 775 for the critical analysis here) can generate parameter estimates and standard errors whose sampling distribution approximates a z-distribution. Because of the difficulty in estimating degrees of freedom for these comparisons, t values are thus treated as approximating z values (Baayen, 2008).

For each segment, we first constructed a baseline model to specify the random effects structure and to identify the significant control variables. Each baseline model included random intercepts for subject and item, as well as random slopes for the contrasts. These slopes were removed if they correlated .9 or greater with either each other or the intercepts, to reduce redundancy and avoid overparameterization. After adding the random effects, each model was built stepwise to include the following control variables: log frequency (based on the original Snodgrass & Vanderwart, 1980, English names), imageability, visual complexity, and familiarity of the experiment target (taken from Rossion & Purtois, 2001), and the trial number (to control for practice effects). Four of these control variables (imageability, visual complexity, familiarity, and log frequency) were sometimes correlated with each other. Any variable that was correlated

with any of the others at .3 or greater was regressed in a model that included all three of the other variables as predictors, and the residuals replaced the original predictor. This provided a "purer" measure of each variable,, while allowing us to test simultaneously for effects of the other variables. The stepwise procedure ensured that only variables which approached significance in this baseline analysis (a *t* value of approximately 1.5) were included in the final models alongside the predictors of interest.

Using the baseline model, we added the predictors of interest, namely a set of contrast variables defined over conditions. Orthogonal contrasts allowed us to make three comparisons: between the both and speaker conditions, between the listener and control condition, and between the combination of both and speaker, and listener and control. The first contrast allowed us to see whether the speaker knowing that the primes were mutual knowledge instead of privileged led to additional reduction. The second tested the opposite: whether knowing that the listener had relevant privileged information led to reduction relative to knowing that no priming information had been provided. Finally, the third tested whether the two pairs differed from each other.

Results

A summary of the segment durations is shown in Table 1. There was a strong effect of the speaker's experience, such that onset and target word durations were shorter when the speaker had heard the prime, i.e. in the Both and Speaker-only conditions. There was no effect of the listener's knowledge on either onset or target durations, in that the Both condition did not differ from the Speaker-only condition, and the Listener-only did not differ from the None condition. The only hint that the speaker was sensitive to addressee knowledge emerged in the determiner region, which was shortest in the Given-to-Both condition, when both speaker and

addressee had heard the prime. The action word was not affected by the manipulation. Tables 2 shows the parameter estimates and overall model designs for each segment of the utterances.

(Tables 1 and 2 about here)

Onset. The parameter estimates for the onset to speak indicated that there was a significant effect of contrast 3 (between Given-to-Both + Speaker-only and Listener-only + Control; β = -.08, SD = .0038, t = -21.1 , p < .0001), with shorter latencies on Both and Speaker trials. Neither the contrast between Both and Speaker-only (β = -.0029, SD = .0052, t = -0.56 , p > .58) nor between Listener-Only and Control (β = -.008, SD = .0053, t = 1.5 , p > .13) was significant.

Determiner. The parameter estimates for the determiner indicated that there was a significant difference between the Given-to-Both and Speaker conditions (β = -.02, SD = .0059, t = -3.37 , p < .0008), with shorter durations in the Both condition. There was also a significant effect of contrast 3 (β = -.011, SD = .0043, t = -2.51 , p < .01), with shorter durations in the Both+Speaker-only combination. The parameter estimate for contrast 2 was not significant (β = -.0044, SD = .006, t = -0.73 , p > .47).

Target word. Most importantly, the model for the target word revealed the same pattern as for the onset latency. Contrast 3 was significant (β = -.013, SD = .0024, t = -5.31 , p < .0001), with shorter durations in the Both+Speaker-only combination. Contrasts 1 (β = -.0015, SD = .0033, t = 0.457 , p > .65) and 2 (β = -.0012, SD = .0034, t = -0.36 , p > .72) were not significant.

Action word. None of the contrasts were significant in predicting the duration of the action word.

Analysis of Speaking Rate. We performed an additional analysis of each region to determine the extent to which our effects were driven by overall speaking rate. For these models,

we included the durations of the other segments in the instruction. The model for object duration, for example, included the onset, article, and action word duration, for example. The test of interest then became whether our original condition contrasts remained significant even after controlling for the effect of speaking rate. Indeed, after including the additional parameters in the model, all of the previously-reported effects remained significant. This holds for all of the models reported in all experiments.

Discussion

The results show a clear effect of speaker-internal facilitation on reduction, as well as a weaker, and somewhat atypical, effect of audience design. Speakers reduced the duration of the object word, and began speaking sooner, when, and only when, they themselves heard the priming information. Knowing that the word was also heard by their addressee had no additional effect on reduction, and no reduction occurred when the speaker did not hear the word but knew the addressee had. Because the object word contains the most task-relevant information, this conflicts with an account where common ground alone drives reduction.

An alternative explanation, that speakers' experience guided reduction, is more consistent with facilitation-based accounts like the FRH. Hearing the prime in this experiment most likely activated representations involved in processing the target word, which both sped planning and smoothed the process of articulation, which as a consequence led to reduction of both latency and object word duration.

The only effect of audience design appeared on the determiner *the*, which shortened when both speaker and listener had heard the prime relative to when only the speaker had heard it. Most accounts of audience design focus on shared referents and/or information and the corresponding referential phrase(s). To our knowledge, no common ground-based account

makes specific predictions about reduction in planning regions due to common ground, although neither do they exclude the possibility.

Understanding the likely mechanism that drives these effects requires a closer look at the conditions and how they influence processing. The Speaker and Both conditions each elicited reduction on the object word and the latency, suggesting that the prime facilitated a significant part of processing in each case. The primary difference between the Speaker and the Both condition, though, lies in the shared information, and the most natural explanation of the shorter *the* in the Both condition is that the speakers recognized the shared givenness, which made the third object fully predictable. One possibility is that speakers attended to the predictability of this referent, and not to the listener *per se*, or to a representation of what she currently knows. Another possibility is that the speaker tracked the addressee's knowledge, which guided her own attention to shared objects, leading to facilitation.

Both of these reflect an audience design-style explanation mediated by the facilitation of a speaker-internal mechanism (cf. Arnold et al., 2012; Horton, 2007). Under this view, the speaker's representations of the listener's knowledge and attention, and/or shared predictability, have the potential to guide the speaker's attention and lead to planning facilitation. Of particular interest is that the common ground effect here appeared only in the planning region, i.e. the duration of the determiner *the*. This echoes the effect of listener attention on determiner reduction that was reported by Arnold et al., 2012. However, we are cautious in drawing strong conclusions about this effect, because the difference in the determiner duration between Both and Speaker conditions was extremely small (avg. 7 msec difference), leading to the possibility that the effect is simply spurious. Further research is necessary to fully understand this pattern.

Proponents of common ground might argue that this situation does not constitute an adequate test of speakers' use of common ground, because not even in the Both condition do the participants explicitly establish their shared representations *as* shared. There is extensive evidence that common ground is used most strongly when it is grounded, i.e. when the addressee provides feedback that the information is shared (e.g. Brown-Schmidt, 2012). For example, comprehension studies have demonstrated that overhearers (i.e. non-participants in a conversation) have more difficulty interpreting utterances than addressees (Schober & Clark, 1989; Wilkes-Gibbes & Clark, 1992).

We acknowledge that a different or more interactive manipulation of common ground may have produced a stronger effect (cf. Brown-Schmidt, 2009). However, it is important to note that our experiment provided numerous opportunities for the speaker to track the listener's knowledge. The speaker saw icons characterizing the listener's knowledge, and both participants experienced the primes in explicitly-delineated blocks. The fact that we did not find robust effects of common ground means that the constraints on reduction are extremely restricted. This would be at odds with the proposal that common ground is the primary constraint on acoustic reduction. This view would also contrast with claims that interlocutors assess common ground on a wide variety of information, including community membership and linguistic co-presence (Clark, 1996; Clark & Marshall, 1981).

But perhaps more importantly, the results display a division based on speakers' experience with the primes. Removing common ground as an explanatory factor leaves this division unexplained, and facilitation accounts provide a ready alternative. If so, this also leads to specific predictions about how the speaker's experience leads to reduction, in particular her experience articulating a word.

Experiments 2a and 2b

Experiment 1 shows that simply hearing the prime, even privately spoken by a third party, suffices to elicit reduction. We argue that this is best explained by facilitation-based accounts, which focus on the speaker's experience. In spoken dialogue, one process that distinguishes the speaker's and addressee's experience is articulation. Both common ground and facilitation accounts predict that information that is given by a third party should elicit reduction, but they make different predictions about the speaker's articulatory experience. Common ground accounts suggest that it should not lead to any additional reduction, in comparison with a situation where all interlocutors hear the word. By contrast, the FRH predicts that speakers may reduce a word as a result of having actually articulated the word recently, independently of the facilitation that comes from the lexical and phonological representations associated with the word. Experiments 2a and 2b are designed to test this hypothesis while separately deconfounding two factors: speaker and listener experience, and saying versus hearing a prime.

Given the results of Experiment 1, Experiment 2a does not include a listener, to test whether articulation alone suffices to elicit reduction. We isolate the articulatory level of production here by asking participants to name prime objects either aloud or silently to themselves. To the extent that these two types of priming differentially activate the articulatory mechanisms, we expect more reduction after having spoken the prime aloud.

The distinction between inner and overt speech finds support in recent studies on tongue-twisters (e.g. Oppenheim & Dell, 2010). These demonstrate that unmouthed inner speech (i.e. speaking silently to one's self) does not exhibit the same pattern of phonemic errors as mouthed inner speech (i.e. silently mouthing words) or overt speech. Inner speech does, however, exhibit

the same lexical bias as overt speech, suggesting that it activates higher-level processes such as lexical selection.

Experiment 2a examines how speakers pronounce words that they have either named aloud before (i.e., those they have articulated) or merely prepared to say sub-vocally. This experiment couples articulation with hearing the prime spoken aloud. A difference in reduction between naming aloud and naming silently may thus arise not because of articulatory facilitation but because the speaker has heard the object named. Experiment 2b pairs speakers with listeners and manipulates which participant makes the prime information part of the discourse. In this case, speakers hear primes in both cases. After speaking themselves, however, speakers should have their articulatory mechanisms facilitated. If this facilitation contributes to reduction, we should observe more reduction after the speaker contributes than after the listener does.

Together, Experiments 2a and 2b test the prediction that the articulatory level of production makes a contribution to reduction. 2a isolates the effect of the speaker from any effect of the presence of a listener, and 2b reintroduces the listener to test the difference between merely hearing a prime and both hearing and speaking it.

In order to focus on articulatory facilitation, Experiments 2a and 2b used designs that reduced the potential influence of target predictability. This differed from Experiment 1, in which the prime made the target object fully (100%) predictable, (cf. the discussion of 'referential predictability' in Kahn & Arnold, 2012; Lam & Watson, 2010; Watson, Arnold & Tanenhaus, 2008). In Experiments 2a and 2b, all trials contained a prime, but it was only the same as the target object on a subset of trials (33% in Exp. 2a, 50% in Exp. 2b). Speakers thus could not make reliable predictions about the relationship between the prime and the target.

Experiment 2a

Method

Participants. A total of 20 undergraduates participated in an instruction-giving experiment as speakers for course credit.

Materials and Design. The same objects were used as in Experiment 1. 24 of these objects were termed experimental targets; these served as the target object in experimental trials, and were chosen to have names with two syllables and similar values on each of the normed dimensions, as well as frequency. The remaining objects served as fillers.

On each trial, 8 objects appeared on screen, including one we will call the **prime object**, and one we will call the **target object**. After 2 seconds, the prime object would flash twice, and a prompt would appear on the screen, telling the participant to either name the prime object out loud, or to name it silently. After another second, the participant could press a button indicating completion of the naming, at which point the target object would perform an action. The participant then had to describe the action, just as in Experiment 1. Clicking on the target object after describing the action advanced the trial.

There were five trial types, including four experimental types, and one set of fillers. The experimental trials formed a 2x2 combination of the two factors of interest (spoken aloud vs. silently, congruent vs. incongruent). On congruent trials, the prime object and target object were the same, whereas on incongruent trials they differed. All other objects for a trial (7 in congruent, 6 in incongruent) were randomly chosen. On spoken aloud trials, the prompt that appeared after the prime object flashed told the participant to name the object aloud. The prompt on silent trials asked them to name the object silently to themselves. The 24 experimental targets appeared once in each of these experimental conditions throughout the experiment, with two constraints: no

item could appear on two trials in a row, and no trial type could occur two trials in a row. All filler trials had different prime and target objects, both of which were randomly chosen. These appeared twice as often as other trial types, for a total of 144 trials (96 experimental (24 in each cell), and 48 filler).

Procedure. The participants each sat in front of a computer, were instructed that they would see arrays of objects, and were given an example. The speaker was told that he would be giving instructions about the movement of these objects, and that the utterances would be recorded for use in a future experiment. Examples of the possible actions (expand, rotate, shrink, and fade) were then shown with randomly chosen objects, and a sentence frame was provided. For example, the participant might have seen an airplane rotating, and would have been told "You might see this [the airplane rotating], and you might say 'The airplane rotates.'" 10 practice trials followed these instructions, all fillers. Participants were allowed to take a break at the halfway mark.

Analysis. The acoustic analyses were identical to those performed in Experiment 1. The statistical analysis also followed a similar procedure, with the exception that two binary contrasts compared the spoken-aloud versus silently-named conditions, and the congruent versus incongruent conditions. The models also included the interaction of these contrasts when significant.

Results

A summary of the segment durations is shown in Table 3. The critical finding concerned the target words. Targets were shorter when primed with the same word, and also in the spoken-aloud condition overall. But targets, determiners, and action words were shortest when the prime

was the same word and was also spoken aloud. Table 4 shows the parameter estimates and overall model designs for each region. We report the results in that same order.

(Tables 3 and 4 about here)

Onset to speak. The parameter estimates for the onset to speak indicated that there was a significant effect of congruence (β = -.035, SD = .0053, t = -6.61, p < .0001), with shorter latencies in response to congruent trials. No other significant effects emerged.

Determiner. The parameter estimates for the determiner showed a different pattern, with significantly shorter durations in spoken-aloud trials (β = -.024, SD = .0071, t = -3.35, p < .0009), and a significant interaction between prime type and congruence (β = -.028, SD = .013, t = -2.18, p < .03), where determiners were shortest in the congruent and spoken-aloud condition. The estimate for congruence showed a non-significant effect of shorter durations for congruent trials (β = -.005, SD = .0066, t = -0.76, p > .45).

Target word. The parameter estimates for the target word revealed two main effects and an interaction. The effect of prime type was significant (β = -.017, SD = .0031, t = -5.46, p < .0001), with shorter durations for spoken-aloud trials. The effect of congruence was also significant (β = -.014, SD = .0028, t = -4.81 , p < .0001), with shorter durations for congruent trials. Finally, the interaction between prime type and congruence was also significant (β = -.02, SD = .0055, t = -3.29, p < .001), indicating that congruence and spoken-aloud combined to produce even shorter durations than the effects by themselves.

Action word. The parameter estimates for the action word displayed only an interaction between prime type and congruence (β = -.016, SD = .0059, t = -2.72, p < .007), again indicating that words were shortest when the prime was both congruent and spoken aloud. The other effects did not approach significance.

Discussion

The results demonstrate a clear effect of the articulatory level of processing. The determiner and object regions both exhibited effects of congruence, but also importantly an interaction with speaking aloud. The congruence effect confirms that the manipulation successfully primed speakers' representations, similar to the results from Experiment 1. The additional interaction effect suggests that over and above congruence, speakers were facilitated by naming the object aloud.

The pattern of results supports the prediction that more facilitation across levels of processing should lead to more reduction. The effect on the onset to speak suggests that having easier or earlier access to the object speeds production, even in the absence of strong predictability. Having spoken the word aloud also appears to speed the articulation of the phrase, including both the determiner and the target word, and the action word when congruence and speaking aloud come together. The presence of an effect even in the absence of a listener suggests that speakers either reduce simply because of their own representations, or maintain a representation of an imagined listener.

Experiment 2b

The most important finding to emerge from Experiment 2a is that speakers appear to have reduced based on having facilitated articulators. Another possible explanation of the results, however, is that speakers react to hearing the objects named out loud. Experiment 2b reintroduces the listener in an attempt to rule out this alternative. The FRH predicts that speakers should reduce more after introducing a prime themselves than after hearing their listener introduce it, because of their facilitated articulators.

Method

Participants. A total of 26 pairs of undergraduates participated in an instruction-giving experiment as speakers for course credit. Four pairs were excluded for either technical difficulties or non-compliance, for a total of 22 pairs of participants.

Materials and Design. The same base set of materials was used as in the first two experiments. Here, we used 96 objects, divided into four sets. The items in each set were chosen to make each set's average frequency, number of syllables, and norming ratings approximately equal. Item sets were rotated between participants, such that each participant pair saw all 96 items, but each item appeared in only one of four conditions for each participant. Across all participants, each item set appeared in each condition approximately the same number of times.

As in Experiment 1, on each trial 8 objects appeared on both participants' screens. However, in this experiment, each trial contained two instructions: a prime instruction (which was always a flash), and a target instruction. The same participant always gave instructions for the target action, this participant is called "speaker". The other is called "listener" (even though listeners also spoke on half of the prime instructions). The prime instruction was initiated when the prime object flashed on either the speaker or listener's screen. If the speaker saw the airplane flash, for example, she told the listener "Make the airplane flash", and vice versa for the listener. After the prime object instruction had been executed, the target object moved, always on the speaker's screen. The speaker then instructed the listener to perform the action, just as in Experiment 1.

The participants were instructed to use the carrier phrase "Make the [object] [action]." The change in carrier phrase from Exps. 1 and 2a was motivated by concerns that participants tend to deaccent words in subject position less often than words in object position (Hirschberg &

Terken, 1993). The new phrase thus increased the chances of detecting an effect. The task was also slightly more interactive than the previous two experiments, in that it required active listener participation.

There were four trial types, resulting from the cross between prime congruence (congruent with target vs. incongruent) and the person who spoke the prime (speaker-given vs. listener-given). Thus in half the trials, speakers spoke the prime, while in the other half listeners did. Similarly, half the time the prime and target object were the same, and half the time they were different.

Procedure. The instructions were very similar to Experiment 2a's. The participants were told they would see the same array of objects, and then one of those objects would flash, either on the speaker or listener's screen. Whoever saw the object flash was to tell the other participant "Make the [object] flash." The recipient of this instruction would then click on the Flash button, followed by the object. Participants were then told that after the flash instruction had been executed, an object would move on the speaker's screen. The speaker was to issue an instruction to the listener with the same sentence frame as the flash instruction, i.e. "Make the [object] rotate/shrink/expand/fade". Trial order was randomized, such that no trial type could appear more than twice in a row. Participants were thus asked to pay attention so that they didn't miss the flashed object. They were allowed to take a short break at the halfway mark (48 trials), but most participant pairs simply continued until the end of the experiment.

Analysis. The acoustic analysis was identical to the ones performed in Experiments 1 and 2a. The statistical analysis took the same form as Experiment 2a, with a 2x2 binary contrast between congruent versus incongruent, and speaker- versus listener-directed priming information.

Results

A summary of the segment durations is shown in Table 5. Congruence, whether from the speaker or listener, produced shorter durations in every region. The object word durations did not differ by speaker- versus listener-givenness, but the determiner durations were shorter in the speaker-given congruent case than listener-given congruent. Table 6 shows the parameter estimates and overall model designs for each segment.

(Tables 5 and 6 about here)

Onset to speak. The parameter estimates for the onset to speak indicated that there was a significant effect of congruence (β = -.045, SD = .0096, t = -4.73, p < .0001), with shorter latencies in response to congruent trials. Neither the effect of speaker givenness nor its interaction with congruence was significant.

Make. The parameter estimates for the duration of the word *Make* showed the same pattern, with a significant effect of congruence (β = -.025, SD = .004, t = -6.22, p < .0001) and no other significant effects.

Determiner. The parameter estimates for the determiner showed a different pattern, with significantly shorter durations in congruent trials (β = -.035, SD = .0088, t = -4.03, p < .0001), and a significant interaction between speaker givenness and congruence (β = -.029, SD = .012, t = -2.36, p < .02). The estimate for speaker givenness was not significant (β = -.007, SD = .0067, t = -1.06, p > .29).

Target word. The parameter estimates for the target word revealed a significant effect of congruence (β = -.053, SD = .0043, t = -12.21, p < .0001). The effect of speaker givenness and its interaction with congruence did not approach significance.

Action word. The parameter estimates for the action word also revealed an effect of congruence (β = -.014, SD = .0047, t = -3.02, p < .003), but again, the other effects were not significant.

Discussion

The main finding from Experiment 2b is that speakers produced mentioned target words with shorter duration than unmentioned target words, and they did so equally when the word was speaker-contributed as when it was listener-contributed. This effect was also reflected in all other regions. By contrast, the determiner duration was the only region where durations were shorter when the speaker had both said and heard the word, as opposed to just heard it mentioned. Critically, this difference was not observed for the target word, which contrasted with those of Experiment 2a.

Importantly, this finding is at odds with the specific prediction of the FRH that articulatory facilitation should lead to greater reduction on the target after the word was introduced by the speaker, as opposed to by the listener. This raises questions about why articulatory facilitation was observed in Experiment 2a, but not here. The major difference between these experiments was that Exp. 2a compared a spoken-silently with a spoken-aloud condition, whereas Exp. 2a had two spoken-aloud conditions. There are two possible interpretations of this difference. The common ground view suggests that these findings stem from a need for speakers to have information available in common ground to have it trigger reduction, because only the spoken-aloud prime can be considered part of common ground. However, we reject this account, because Exp. 1 demonstrated that common ground is not a requirement for reduction (see also Bard & Aylett, 2004). The other possibility is that although

articulatory facilitation may have a weak effect, hearing the word spoken aloud also provides a strong facilitatory effect itself, which masks any additional facilitation from speaking the word.

Support for this second view comes from the observation that we did observe a small difference between speaker-given and addressee-given conditions, again on the duration of the determiner. This emerged as a significant interaction between congruence and speaker givenness. An examination of the means suggests that in fact, speakers are producing the longest determiners when they see an incongruent prime, and exhibit greater reduction when they speak the prime than when the listener does. This pattern is consistent with the claim that facilitated production processing leads to reduction. The primary difference between the speaker-given and listener-given trials is that in the speaker-given trials, the speaker must wait an indeterminate amount of time after issuing the priming instruction for the listener to execute. This makes the movement of the target object difficult to predict, especially on the incongruent trials. The difference between the speaker-given congruent and speaker-given incongruent trials could plausibly be attributed to speakers directing their attention more quickly and/or easily in the congruent trials, thereby easing the processing of the planning regions. This effect is thus explainable as a speaker-internal effect, consistent with the findings from Arnold et al., (2012).

The fact that this effect did not extend to the target word may be attributable to the potential facilitatory effects of hearing the word. As shown in Experiment 1, speaking the word is not necessary for reduction. This could be explained by the idea that facilitation occurred for the common representations used for comprehending and speaking, such as the concept and lemma representations. In addition, there is precedent for the idea that the articulatory mechanisms themselves may be primed by hearing a word. The Motor Theory of Speech Perception (for a recent review, see Galantucci, Fowler & Turvey, 2006) suggests that in order to

perceive an utterance, listeners make use of the motor codes they would use to produce that utterance themselves. When applied to a facilitation account of reduction, this would predict that hearing someone else produce the word led to facilitating motor processing, and thus to reduction.

In sum, the results of Experiments 2a and 2b suggest that hearing a word has strong effects on reduction – whether the word was spoken by oneself or by another, or in the presence of an addressee at all. By contrast, if articulatory facilitation supports reduction, it does not do so strongly enough to appear in the context of other sources of facilitation, such as hearing the word. In addition, we again demonstrated that facilitatory effects may show up on the determiner, and not just the target noun itself, supporting accounts in which reduction is related to the facilitation of production planning.

General Discussion

The results reported here demonstrate that acoustic reduction is influenced by variations in the speaker's own prior experience with the information being articulated. In Experiment 1, onset latencies and word durations were shorter when the speaker had heard the prime, regardless of what the listener had heard. In Experiment 2a, onset latencies and word durations were shorter when the speaker had said the prime out loud than said it internally. In Experiment 2b, onset latencies and word durations were shorter when the speaker had either said or heard the prime than when they hadn't.

The findings reported here contribute both empirically and theoretically to the literatures on acoustic reduction, common ground, and speech production. Empirically, they demonstrate that durational reduction is affected by the speaker's experience of hearing or speaking a prime, regardless of the presence, knowledge, or participation of the addressee. They also show that

these effects extend to multiple regions in the utterance, including both the target and planning regions. Some weaker manipulations even emerged only in the planning regions, and not on the target word.

Theoretically, the current findings place constraints on the kinds of psycholinguistic models that could explain reduction. We have considered two theoretical approaches here: common ground and facilitation approaches. Our data suggest that the strongest versions of both approaches must be reconsidered, but generally support a facilitation approach.

On the one hand, our results do not support theories in which acoustic reduction is primarily the result of audience design (e.g. Lindblom, 1990). In Experiment 1, speakers reduced after primes in the Speaker-only condition, which would be infelicitous from a common ground perspective. Moreover, we did not see any increased reduction in the Both condition. These findings add to a series of studies that have looked for common ground effects on acoustic reduction and failed to find them (Bard et al., 2000; Bard & Aylett, 2004). We do not, however, conclude that common ground has no effect on acoustic reduction, given that other studies have found that acoustics do vary according to audience knowledge or behavior (Arnold et al., 2012; Galati & Brennan, 2010; Rosa et al, under review, this volume). In fact, even in this study, we found a hint of audience design, in that speakers in Experiment 1 used shorter determiners when the target prime was shared. Below we consider the possibility that these effects may also be related to production facilitation.

On the other hand, our results also suggest that one prediction of the FRH was overly strong, namely the claim that any facilitation should lead to reduction. In particular, our position predicts that speakers should have reduced more after speaking the prime than after merely hearing it in Experiment 2b, similar to their performance in Experiment 2a. Two interpretations

are possible. The first is that together, Experiments 2a and b show that *hearing* the prime plays the most important role in creating durational reduction. The second is that the greater interactivity of the task in Experiment 2b (as opposed to Experiment 1) created floor effects, such that it elicited "maximal" reduction in both the speaker- and listener-given conditions, producing no difference.

Nevertheless, the results presented here provide good overall support for accounts of reduction that focus on speaker experience. This conclusion stems from four observations. First, reduction in Experiment 1 occurred only when the speaker had heard the prime. Second, Experiment 2a revealed greater reduction when the speakers had actually said the prime than when they had just produced it subvocally. This may have resulted from the activation of articulatory processes, or from having heard the word, either of which is likely to have facilitated later productions. Third, in Experiment 2b there was a hint of a difference between speaking and hearing the prime, such that the determiner was reduced even more when the speaker had said the prime than when the listener had. Fourth, the priming manipulations in all experiments tended to result in reduction over multiple regions, and not just the target word itself. For the most part, when a predictor had an effect on one region, it affected three regions: the onset to speak, the determiner, and the target word (with the exception of Exp. 2a, in which onset times were not specifically affected). This effect persisted despite all models including the durations of the other utterance regions as predictors, and is thus not a simple one of speech rate.

Notably, the effects in non-target regions occurred primarily to the regions in which the speaker was likely planning the target word, i.e. the latency to speak and the determiner. This pattern of results suggests that previous mention and articulatory facilitation affect the planning and articulation of the target word, as distributed over multiple words in an utterance. This

provides the strongest support for facilitation-based accounts, because they predict that reduction should appear anywhere the speaker experiences ease of processing, including regions of an utterance that do not convey any particularly relevant information. Other findings are consistent with evidence that even function words and connectives, which carry little information, exhibit reduction in circumstances that make them easy to process (e.g. Bell et al., 2009, Jaeger, 2006).

Of particular interest is the possibility that ease of planning might account for some of the common ground effects observed here and elsewhere. Speakers in Experiment 1 produced shorter determiners when both participants heard the prime than when only the speaker heard it, although only for the third (fully predictable) instruction. The indication that speakers recognized the commonality of the representation, even in a situation where the task made the reference 100% predictable, suggests that speakers track some aspects of common ground. We observed a similar effect in another study (Arnold et al., 2012), in which the determiner was shorter when the listener had already anticipated the target object by picking it up. Common ground may thus guide speakers' attention toward elements of the conversation relevant to processing, and thus to reduction.

The onset and determiner effects also cohere with recent findings from corpus analyses that have been used to support the principle of Uniform Information Density, which holds that speakers attempt to achieve maximal efficiency in communication by smoothly communicating bits of information over time (e.g. Jaeger, 2010). The FRH and other facilitation accounts are consistent with this proposal, but make predictions at a different granularity. Information density patterns hold over large corpora of data, and similarly the claims made about it apply to speakers and utterances on average. No specific claims are made about moment-by-moment calculations of information in particular situations, nor about the mechanisms that might underlie such

calculations. Facilitation accounts, by contrast, make predictions about particular speakers in particular utterance contexts, and the FRH proposes a mechanism by which facilitation might lead to reduction. Thus, facilitation mechanisms might explain why speakers end up producing information at a constant rate, because the contexts in which production is facilitated are those in which the context decreases the informativeness of the word.

In sum, this paper programmatically contrasts specific predictions of the FRH and other similar facilitation accounts with an audience design account of reduction. Moreover, it presents a thorough analysis of participants' utterances, including not just the target words, but also other words in the utterance. This contributes to our claim that processing factors play a role in explaining acoustic reduction phenomena, in addition to categorical, high-level representations like common ground (Arnold & Watson, under review, this volume; Kahn & Arnold, 2012). We argue that a complete understanding of the variation between acoustically prominent and reduced forms requires an understanding of these processing factors.

References

Ariel, M. (1990). Accessing noun–phrase antecedents. London: Routledge.

Arnold, J.E., Kahn, J.M. & Pancani, G. (2012) Audience Design Affects Acoustic Reduction Via Production Facilitation. Psychological Bulletin & Review, 19, 505-512

Arnold, J.E. & Watson, D. (under review, this volume). Synthesizing meaning and processing approaches to prosody: performance matters.

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* (1), 31-56.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge University Press.

Balota, D.A., Boland, J.E. & Shields, L.W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28(1)*, 14-36.

Bard, E.G., Anderson A.H., Aylett, M., Doherty-Sneddon, G., Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*(1), 1-22. doi: 10.1006/jmla.1999.2667

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

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*, 92-111.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89,* 1–47.

Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.15) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(6),482-1493. doi: 10.1.1.121.3930

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893-900. doi:10.3758/PBR.16.5.893

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes 27(1)*, 62-89.

Bybee, Joan. 1999. Use impacts morphological representation. *Behavioral and Brain Sciences* 22: 1016-1017.

Chafe, W. (1976) Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. (ed.). *Subject and Topic*. New York: Academic Press.

Clark, H. H. & Haviland, S.E. (1977). Comprehension and the given-new contract. In Freedle, Roy O. (Ed.), *Discourse production and comprehension.* (1-40) Norwood, NJ: Ablex

Clark, H.H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L. & Sag, I.S., (Eds.). *Elements of Discourse Understanding* (pp. 10-63). New York: Cambridge University Press.

Clark, H.H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37(3)*, 201-242.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* , 22, 1-39.

Clark, H.H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*(3), 201-242. doi:10.1006/cogp.1998.0693

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

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.

Gahl, S. (2008). *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84(3)*, 474-496.

Galantucci, B., Fowler, C.A. & Turvey, M.T. (2006). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review, 13(3).* 361-377.

Galati, A. & Brennan, S.E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*(1), 35-51. doi:10.1016/j.jml.2009.09.002

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177-220). London: Academic Press.

Gorman, K. S., Gegg-Harrison, W., Marsh, C.R. & Tanenhaus, M.K. (in press). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Grosz, B.J. & Sidner, C.L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12(3)*. 175-204.

Gundel, J., N. Hedberg & R. Zacharski (1993) Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274-307. doi:10.2307/416535

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. Mouton: The Hague.

Hanna, J.E. & Tanenhaus, M.K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28(1)*, 105-115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43-61.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*, 831-836.

Hirschberg, J. & Pierrehumbert, J.B. (1986). The intonational structuring of discourse. *Proceedings of the Association for Computational Linguistics*. 136-144.

Horton, W.S. (2007). The influence of partner-specific associations on language production: Evidence from picture-naming. *Language and Cognitive Processes, 22(7).* 1114-1139.

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1-35.

Jaeger, T. F. (2006). Redundancy and syntactic reduction in spontaneous speech. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Jaeger, T.F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61(1)*, 23-62.

Kahn, J.M. & Arnold, J. E. (2012) A Processing-Centered Look at the Contribution of Givenness to Durational Reduction. *Journal of Memory & Language, 67*, 311-325

Lam, T.Q. & Watson, D.G. (2010). Repetition is easy: Why Repeated Referents Have Reduced Prominence. *Memory & Cognition 38*(8), 1137-1146. doi: 10.3758/MC.38.8.1137

Levelt, W. J. (1989). *Speaking: From intention to articulation.* Cambridge, MA. MIT Press.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer Academic.

Prince, E. (1981). Toward a taxonomy of given-new information. In Cole, P. (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thomspon & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295-325). Philadelphia: John Benjamins B.V.

Rosa, E.C., Finch, K., Bergeson, M. & Arnold, J.E. (under review, this volume). The Effects of Addressee Attention on Prosodic Prominence.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition [On-line]. *Journal of Vision, 1*, 413a.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

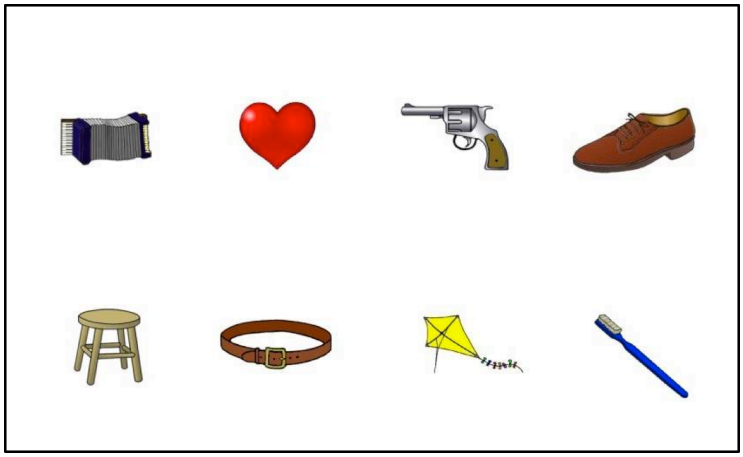
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Venditti, J. J., & Hirschberg, J. (2003). *Intonation and discourse. Invited panel presentation*, *Proceedings of the XVth International Congress of Phonetic Science* (pp. 107-114). Barcelona.

Watson, D.G., Arnold, J.E., & Tanenhaus, M.K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*, 156-1557. doi:10.1016/j.cognition.2007.06.009

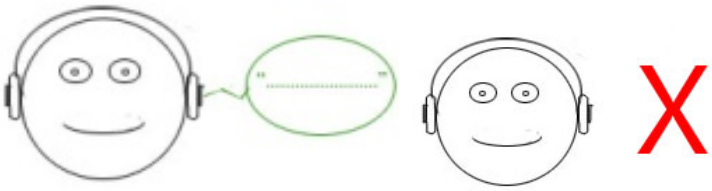
Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183-194.

Figure 1



An example object array.

Figure 2



The icons that the speaker saw in Experiment 2, indicating whether or not the listener heard the priming information. The left icon indicated that the listener heard the prime, the right that he/she did not.

Table 1

Means and standard deviations (in parentheses) from Experiment 1 in milliseconds, broken down by condition and segment.

	Onset	*The*	Target	Action
Both	909 (345)	66 (26)	428 (122)	744 (160)
Speaker-Only	940 (324)	72 (27)	429 (118)	753 (181)
Listener-Only	1396 (603)	75 (36)	448 (113)	763 (175)
Neither	1315 (489)	73 (31)	455 (120)	770 (179)

Table 2

T-*values of the parameter estimates associated with particular variables in the analyses reported for Experiment 1. The columns refer to models of individual segments, and the rows to variables in the analyses. "--" indicates that the variable was not included in the final model for that segment.*

	Latency to Speak	Determiner (*the*)	Target word	Action word
Trial	*t* = -5.73	--	*t* = -1.83	--
Imageability	--	--	--	--
Familiarity	--	--	--	--
Visual Complexity	--	*t* = 1.04	--	--
Log Frequency	--	--	*t* = -3.72	--
Log Onset Duration	--	%%	%%	%%
Log *the* Duration	*t* = 5.51	--	*t* = 3.42	*t* = 2.05
Log Object Duration	*t* = 3.82	*t* = 2.87	--	*t* = 5.56
Log Action Duration	*t* = -8.84	*t* = 1.83	*t* = 6.24	--
Contrast 1 (Both vs. Speaker-only)	*t* = -0.56	*t* = -3.37	*t* = 0.46	*t* = 0.03
Contrast 2 (Listener-only vs. Control)	*t* = 1.5	*t* = 0.73	*t* = -0.36	*t* = -0.32
Contrast 3 (Both + Speaker-only vs. Listener-only + Control)	*t* = -21.1	*t* = -2.51	*t* = -5.31	*t* = -0.45

Table 3

Means and standard deviations (in parentheses) from Experiment 2a in milliseconds, broken down by condition and segment.

		Onset	*The*	Target	Action
Congruent	Aloud	1868 (696)	76 (28)	428 (86)	681 (117)
	Internal	1875 (664)	83 (31)	457 (86)	704 (123)
Incongruent	Aloud	1979 (652)	80 (28)	450 (85)	702 (119)
	Internal	1998 (631)	82 (33)	459 (84)	694 (124)

Table 4

T-*values of the parameter estimates associated with particular variables in the analyses reported for Experiment 2a. The columns refer to models of individual segments, and the rows to variables in the analyses. "--" indicates that the variable was not included in the final model for that segment.*

	Latency to Speak	Determiner (*the*)	Target word	Action word
Trial	t = -18.39	t = -3.84	t = -1.75	t = -3.93
Imageability	t = 1.73	--	--	--
Familiarity	t = 1.4	--	--	--
Visual Complexity	--	--	--	--
Log Frequency	--	--	t = -1.29	--
Log Onset Duration	--	t = 3.7	--	--
Log *the* Duration	t = 3.1	--	--	--
Log Object Duration	t = 2.2	--	--	t = 2.5
Log Action Duration	t = -12.89	--	t = -3.79	--
Spoken Aloud vs. Silently Named	--	t = -3.41	t = -5.64	t = -1.13
Congruent vs. Incongruent	t = -6.6	t = -1.31	t = -5.06	t = -1.13
SASN*CI	--	t = -2.16	t = -3.28	t = -2.72

Table 5

Means and standard deviations (in parentheses) from Experiment 2b in milliseconds, broken down by condition and segment.

		Onset	*Make*	*The*	Target	Action
Congruent	Speaker	1506 (487)	233 (66)	106 (44)	388 (118)	681 (120)
	Listener	1476 (508)	230 (74)	105 (40)	380 (106)	704 (154)
Incongruent	Speaker	1669 (548)	244 (65)	121 (57)	437 (116)	702 (170)
	Listener	1685 (558)	241 (67)	114 (51)	439 (128)	694 (146)

Table 6

T-*values of the parameter estimates associated with particular variables in the analyses reported for Experiment 2b. The columns refer to models of individual segments, and the rows to variables in the analyses. "--" indicates that the variable was not included in the final model for that segment.*

	Latency to Speak	Determiner (*the*)	Target word	Action word
Trial	t = -13.17	t = -3.09	t = -1.03	t = -3.81
Imageability	--	--	--	--
Familiarity	--	--	--	--
Visual Complexity	t = -1.59	--	--	--
Log Frequency	--	--	t = -6.03	--
Log Onset Duration	--	t = 2.81	t = 4.12	--
Log *Make* Duration	t = 4.9	t = 4.54	t = 7.26	t = 3.54
Log *the* Duration	t = 2.78	--	t = 3.42	t = -9.75
Log Object Duration	t = 3.26	--	--	t = -1.7
Log Action Duration	t = -9.61	t = 2.91	--	--
Speaker vs. Listener	t = 0.64	t = 1.05	t = 0.22	t = 0.48
Congruent vs. Incongruent	t = -4.73	t = -3.97	t = -11.68	t = -4.31

S/L * C/I	$t = 0.84$	$t = -2.37$	$t = 0.92$	$t = -0.89$