Effects of Discourse Status and Planning Difficulty on Acoustic Variation

Lap-Ching Keung

University of North Carolina at Chapel Hill

Spring 2013

A thesis presented to the faculty of the University of North Carolina

at Chapel Hill in partial fulfillment of the requirements for the

Bachelor of Science degree with Honors in Psychology.

Committee Chair: Jennifer E. Arnold, Ph.D.

____

Committee Member: Joseph B. Hopfinger, Ph.D.

Committee Member: Jennifer L. Smith, Ph.D.

Acknowledgements

You would not be reading this today if it had not been for the support and mentorship of my adviser, Dr. Jennifer E. Arnold. To her, I give my sincerest thanks, for her patience and guidance, for her feedback and comments on everything that I submitted to her, and for her encouragement to pursue this honors project, the CUNY conference, and graduate school. Thank you for accepting me as a PSYC 395 student two years ago and treating me like one of your own grad students. Since then, I have gained so much insight into the life of an academic, and I am happy to be pursuing it in the future.

To the other members of my committee, Dr. Joseph B. Hopfinger and Dr. Jennifer L. Smith: Thank you for taking the time to participate in my defense, for being available to answer my questions, and for your comments and feedback.

To the UNC Language Processing Lab: I am grateful for Jason Kahn and Elise Rosa, who gave me advice about my project, participants they no longer needed, and tips about grad school. And many thanks to the RAs who helped me run participants, segment sound files, and code: Michael Chen, Natalie Feingold, Ivy Hauser, Sarah Lineberry, Ingrid Tablazon, and Luke Miller.

Last, but certainly not least, to my friends and family: Thanks for listening to the sound files with me, and for suffering with me while I textgridded everything. But more than that, thanks for just being there for me. A special thanks to my lovely parents Yikong Keung and Amy Cheng, for supporting and inspiring me to pursue my interests. Your encouragement throughout my life is the reason I was able to accomplish my goals here at UNC Chapel Hill.

This project was partially supported by (a) the Tom and Elizabeth Long Excellence Fund for Honors, administered by Honors Carolina, and (b) the Department of Psychology Undergraduate Research Grant.

Abstract

Repeated words within a discourse tend to be acoustically reduced, i.e., shorter in duration. This variation can be explained by a pragmatic selection rule (discourse status) or by speaker facilitation (planning difficulty). The main question of interest: Are these effects part of the same cognitive system or different systems? During the experiment, speakers saw an array of four objects and described a sequence of two movements like in (1).

(1) The chiddle moved above the hamel.
The cammer moved above the neeken.

The target's discourse status (new vs. given) and planning difficulty (novel vs. familiar) were manipulated. For novel targets, the combination of abstract drawings plus nonword labels made planning more difficult. At the onset latency, determiner, and target noun, there were main effects of discourse status and planning difficulty. Durations were longer for new vs. given and novel vs. familiar. These results replicated findings from previous research. More importantly, at the target noun, there was a trend towards an interaction between discourse status and planning difficulty. Givenness had a greater effect on novel targets. Although not significant, onset latency patterned in the same way as target duration, showing a correlation between the two factors. These results together suggest that discourse status is partially mediated by planning difficulty, which further suggests that the two effects are operating from the same cognitive system.

Effects of Discourse Status and Planning Difficulty on Acoustic Variation

Depending on the context, speakers vary how they say their words with respect to prominence—this is *acoustic variation*. Sometimes, a word will be *prominent*, which is characterized by a longer duration, more extreme pitch contour, higher intensity, and greater intelligibility than normal. Other times, if not already replaced by a pronoun, it will become *reduced*. It will be shorter in duration, less varied in pitch, lower in intensity, and otherwise less intelligible than normal (Kahn & Arnold, 2012; Kahn & Arnold, under review; Watson, 2010). Here, we ask how durations in particular will be influenced by the interaction of discourse status (e.g., repeated mention) and planning difficulty (e.g., object novelty). To what extent will these two factors influence the degree of durational reduction?

Factors Affecting Acoustic Variation

An ongoing goal in psycholinguistics is to understand the underlying cognitive mechanisms that drive acoustic variation. Why does speech vary the way that it does? Two approaches have been proposed to explain this phenomenon (as characterized by Arnold & Watson, under review).

The message-based approach and discourse status. The message-based approach reflects linguistic competence at the pragmatic level. The speaker says something a certain way because the grammar selects a degree of acoustic prominence to mark that entity's *discourse status* (Arnold & Watson, under review). In general, the pragmatic selection rule states that *new* information is marked with a *prominent* (i.e., prolonged) form, while *given* information is marked with a *reduced* (i.e., shortened) form (Halliday, 1967). When a speaker introduces something new into the conversation, the discourse status of that entity is new, so the grammar selects a prominent form for the word, and the word will be lengthened. On subsequent mentions, the discourse status changes to given, so a reduced form is now selected, thus resulting in a shortened pronunciation (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Brown, 1983; Fowler& Housum, 1987; Jurafsky, Bell, Gregory, & Raymond, 2001).

Research has shown that previously-mentioned references are reduced compared to their first mention (Bard & Aylett, 2004; Bard, Lowe & Altmann, 1989; Brown, 1983; Fowler, 1988; Fowler & Housum, 1987). For example, in an analysis of scripted monologues, words were shorter when they were spoken a second time (Fowler & Housum, 1987). The same pattern was found in unscripted map task dialogues, where repeated mentions to a landmark were shorter (Bard & Aylett, 2004).

The pragmatic selection rule suggests that this phenomenon is more than a result of simple repetition. It has to do with referring to the same referent, that is, something that is given. In a production experiment by Fowler (1988), participants read paragraphs aloud that contained a target word that was preceded by a previous mention or a homophone of the target word. Only repeating the same word (and thus referring to the same referent) resulted in shorter durations. Simply repeating the articulatory pattern for a word (i.e., homophones, which refer to different referents) was not enough to lead to reduction (Fowler, 1988). This pattern supports the distinction between new and given information, as well as the role of the pragmatic selection rule in acoustic reduction.

An audience-design/common-ground account has been proposed to explain why this occurs. It suggests that a speaker keeps track of the listener's knowledge during the conversation. Information is considered given if it is shared by both the speaker and listener (Gundel, Hedberg, & Zacharaski, 1993; Chafe, 1994). In this case, the speaker will use reduced forms. However, when he wants to introduce new information (i.e., information not in the common ground), he selects a prominent form in order to signal its newness and facilitate the listener's identification of the referent (Baumann & Grice, 2006; Baumann & Hadelich, 2003). This account is debated extensively, with research showing that acoustic variation still occurs even in the absence of the listener (Bard & Aylett, 2004; Kahn & Arnold, 2012). As a result, audience design is not a major determinant of acoustic reduction, so not having a listener as part of the task (as in the current experiment) should not eliminate the effects of repeated mention.

In conclusion, the message-based approach largely explains acoustic variation in terms of pragmatics and discourse-level representations: how recently was a referent just mentioned, and how likely is it to be mentioned again. To a smaller degree, acoustic variation also depends on what the speaker thinks the listener knows. This is contrasted by the facilitation-based approach, which is largely based on speaker-internal processes.

Facilitation-based approach and planning difficulty. The facilitation-based approach reflects linguistic performance because it is systematically related to the speaker's ease of producing an utterance, and not necessarily to pragmatics (Arnold & Watson, under review). These two approaches are consistent because given and accessible information is often what the speaker finds easier to retrieve and produce. Prolonged word durations are correlated with complex tasks and situations (Ferreira & Swets, 2002), while conceptual facilitation leads to shorter durations (Balota, Boland, & Shields, 1989).

Under accounts of incremental planning, when a word is difficult to retrieve (e.g., low frequency), the planning region (i.e., everything before the target word, including pauses) will be lengthened (Griffin, 2003; Meyer, Belke, Hacker, & Mortensen, 2007; Meyer, Roelofs, & Levelt, 2003). In a production experiment by Christodoulou and Arnold (2012), speakers named two objects without pausing (e.g., *hanger windmill*). If word 2 was easy to produce (e.g., high frequency), then the duration of word 1 was short. Speakers can devote more resources to forming the utterance, which allows for faster articulation. On the other hand, if word 2 was hard, then speakers would lengthen word 1, presumably to give themselves more time to retrieve and produce a difficult word. Furthermore, speakers need to devote some of their resources to planning the word, which decreases resources available for utterance formation, which further slows down articulation.

The difficulty of planning a word can be reduced by priming, which triggers pre-activation of the word (Kahn & Arnold, under review). In a production experiment by Kahn and Arnold (under review), speakers saw an array of objects and were asked to describe motions (e.g., *The airplane fades*) to a listener. Before each trial, the speaker or listener would be primed as to which object would move. Speakers were told when the listener received the prime, and vice versa, so the information would be in their common ground. Consistent with the facilitation-based approach, acoustic reduction occurred when the speaker was primed, but not when the listener was primed. This suggests that the speaker did not utilize common ground and selected a reduced form when it was easier for him to produce the word, regardless of the listener's knowledge.

Given words, like primed words, are also easier to produce because there is residual activation from the first mention. Once a word is produced one time, representations in the speaker's production system become activated. This includes representations related to concept (higher-leveled process), lexical choice, and articulation (lower-leveled process). The prior activation, assuming it has not decayed yet, leads to easier and faster retrieval during repeated mentions, thereby allowing the words to have faster articulations (Kahn & Arnold, 2012).

In general, facilitation leads to reduction. An easy word poses no planning difficulty for the speaker, so more cognitive resources can be devoted to forming the sentence. A primed word and repeated word both involve pre-activation of representations in the production system, which results in faster activation and articulation during the target word. However, this leads to a further complication: the acoustic reduction of repeated words can be explained by both the message-based and facilitation-based approaches, but for different reasons (Arnold and Watson, under review). The current experiment will therefore attempt to provide insight into how these reasons are related.

Conclusion. The message-based and facilitation-based approaches are in competition with each other because they postulate different cognitive processes to account for acoustic variation. According to the first one, acoustic variation is a factor of pragmatics and context: “select reduced forms for given referents and prominent forms for new referents”. On the other hand, according to the second one, the variation is instead a factor of speaker-internal processes: “reduce a word if it is easy to produce; otherwise, lengthen it”. As of now, we cannot determine whether one, the other, or both are at work in repeated-word reduction. There is experimental support for both of these mechanisms, but how they relate to each other remains unclear. Do they act separately, or do they work together to affect acoustic variation?

Current Experiment

An open question is whether the effects from the message-based approach (discourse status) can be separated from the effects of the facilitation-based approach (planning difficulty), or whether they interact. Discourse status and planning difficulty may work independently from one another, and if so, these effects will be additive. But if not, then it may predict a potential interaction. Here, we specifically ask if recently learned (novel) and highly practiced (familiar) words display the same effects of repeated mention on acoustic reduction. Assuming only an effect of discourse status, then repeated novel words should be reduced to the same degree as repeated familiar words. But what does it mean if novel words are reduced proportionately more than familiar words? After all, once a novel word becomes given, it also becomes more familiar, solely from the fact that it has already been activated once. If we find such an interaction, then it may suggest that effects from discourse status are partially mediated by planning difficulty (Arnold & Watson, under review). In other words, how much a repeated word is reduced would depend on how difficult that word is to plan.

In this experiment, the novelty of a word is a measure of planning difficulty. The participants see two types of stimuli: familiar (e.g., *camel* and *hammer*) and novel (e.g., *cammer* and *hamel*). These are listed in Appendix A. Novel words are used as labels for abstract drawings, which results in two manipulations of lexical access: frequency and conceptual difficulty. These novel stimuli, which have a frequency of zero and are paired with the abstract drawings, make planning extremely hard.

Word frequency is often considered a measure of planning difficulty, with low-frequency words taking longer to produce (Forster & Chambers, 1973) and being less reduced (Bell et al., 2009; Pluymaekers, Ernestus, & Baayen, 2005). The motivation for using nonwords comes from the observation that while low-frequency words are hard to retrieve and produce, they also tend to have a longer inherent word length (Zipf, 1935). Our paradigm de-confounds this tendency such that the novel words have a frequency of zero, but they are also matched to the familiar words in terms of syllable structure, phonological length, and phonotactic probability. This ensures that any observed differences in durations are the result of novelty (and the discourse status manipulation).

During the experiment, the participants describe a sequence of two moving objects. The second object (target) is manipulated in terms of repeated mention: Nonrepeated targets like in (1) are "new," and repeated mention targets like in (2) are "given". Targets are underlined in these examples. The second time the participants say *cammer* in (2), it is likely to be acoustically reduced compared to the *cammer* in (1).

(1) The chiddle moved above the hamel.

The <u>cammer</u> moved above the neeken.

(2) The cammer moved above the hamel.

The <u>cammer</u> moved above the neeken.

The critical question then becomes whether the reduction in *cammer* is more or less noticeable than the reduction in a familiar word like *camel*. When first mentioned, *cammer* will likely have a longer duration than *camel*. This largely reflects planning difficulty. *Cammer* is a made-up label for an abstract drawing, and it has competitors such as *camel* and *hammer* that are easier to retrieve. Then, once it has been retrieved successfully, the repeated *cammer* may end up having a relatively short duration, possibly even as short as a repeated *camel*. That is, a novel word will be reduced more than a familiar word.

A simple explanation is that there is more room for *cammer* to be reduced (i.e., a longer duration allows for more reduction). It is also possible that the ease of re-retrieving *cammer* becomes comparable to the ease of re-retrieving *camel*. That is, the extra planning difficulty associated with *cammer* should no longer be relevant during the repeated mention—the word was already accessed once and is now stored in working memory for faster access. Either way, the degree of reduction should be greater for *cammer* than it is for *camel*, suggesting an interaction between discourse status and planning difficulty.

The current experiment forms a part of ongoing research on the psychological mechanisms that underlie the human ability to produce language. We focus on discourse status and planning difficulty here because (a) both of these factors are known to result in acoustic reduction independently, and (b) there is debate on how they interact and work together, if at all. We aim to replicate previous findings that repeated-mention words are likely to be acoustically reduced, but we will also attempt to show how repeated mention interacts with planning difficulty. As a result, this study will try to experimentally relate these two factors and contribute to the ongoing debate.

Method

Participants

A total of 24 students and employees from the University of North Carolina at Chapel Hill participated in the experiment. Thirteen of them received course credit, and 11 were paid $7.50 for 45 minutes. All participants were native English speakers, with normal or corrected-to-normal vision and no history of speech impediments.

Materials

A set of 16 familiar and 16 novel word-picture pairs served as the target items. These are listed in Appendix A. All words were disyllabic with the primary stress falling on the first syllable. Neither syllable could be an English word on its own.

Familiar objects. Sixteen colorized versions of the Snodgrass and Vanderwart (1980) line drawings served as the familiar targets (Rossion & Pourtois, 2004). They were chosen because the drawings had similar values for imageability, visual complexity, and familiarity (as reported in Rossion & Pourtois, 2004), and their names had similar frequencies (as retrieved from the Corpus of Contemporary American English; Davies, 2008-).

Novel objects. Sixteen abstract line drawings (from past experiments conducted in the UNC Language Processing Lab) served as the novel targets. Their names were created by recombining the first syllable of one of the familiar words with the second syllable of another familiar word, and vice versa (e.g., *camel* and *hammer* became *cammer* and *hamel*). Having to associate abstract names with abstract drawings made the stimuli sufficiently hard to plan. The following criteria were used when creating these words.

- The two familiar words undergoing recombination must have the same syllable structure. For example, the first syllable of *camel* and *hammer* is a consonant followed by a vowel, and the second syllable is a consonant followed by a syllabic consonant.
- The segments at the syllable boundary must be matched as closely as possible with respect to voicing, place, and/or manner of articulation. For example, these segments are matched on all three dimensions for *camel* and *hammer* (i.e., they are the same: [kæ.ml̩] and [hæ.mr̩]).

These two criteria ensured that the novel words were maximally similar to the familiar words. This was further quantified by two independent phonotactic probability calculators (Hayes, 2012; Vitevitch & Luce, 2004). Phonotactic probability refers to the frequency that a sequence of sounds, such as [æm], occurs in a given position in a word. A two-tailed *t*-test revealed that the scores for the novel words did not significantly differ from the scores of the familiar words, $t(14) = 0.33$, $p = 0.99$ and $t(30) = 0.36$, $p = 0.94$, respectively.

With the stimuli as described above, the phonological components of the familiar and novel words are close to identical, and as such, their average durations should in theory be close to identical. However, processing constraints are expected to prevent that from happening. The participants have never encountered the novel words or abstract drawings before, so these words will require more time and more cognitive resources during both retrieval and production. Therefore, it is likely that novel words will be produced with a longer onset latency and word duration as compared to familiar words.

Design

There were a total of 64 trials, which were divided into two blocks of 32 trials, which were further divided into four sub-blocks of eight trials. The order of the blocks and sub-blocks was counterbalanced.

A training session took place before each sub-block, during which the participant learned the names of the eight objects (four familiar and four novel). Familiar objects never appeared in the same sub-block as their novel counterparts. The participant was instructed not to say any of the words aloud, was allowed unlimited times to listen to them, and was given as much time as necessary to memorize the word-picture pairs before taking a multiple-choice quiz. The quiz was administered again if the participant failed to match two or more of the four novel objects. Then, the experiment proceeded with the actual trials.

The general structure of a trial was as follows. Four objects appeared on the computer screen as in Figure 1; they were either all familiar or all novel. As a result, each set of four objects was seen four times per sub-block, with their positions different each time. After one second, one object would move above another. The participant was explicitly told to use the verb *moved above* in his responses. Thus, the first movement prompted the participant to say something like (3). A second movement then occurred (either the same object moved or a different one), which prompted the participant to say something like (4). This marks the end of one trial. Each sub-block was comprised of eight trials, with eight different objects serving as the target (*cammer* in this example). Each object also served as the alternate (*chiddle*) and both goals (*hamel* and *neeken*) at least once within the sub-block.

(3) The chiddle moved above the hamel.

(4) The <u>cammer</u> moved above the neeken.

Discourse status was also manipulated. In the given condition, the same object would move both times, while in the new condition, two different objects moved. The two blocks reflected this manipulation. The objects that were given in Block 1 would be new in Block 2, and vice versa. With this design, all participants saw all trials, which allowed for within-subject comparisons, assuming no significant difference between the two blocks.

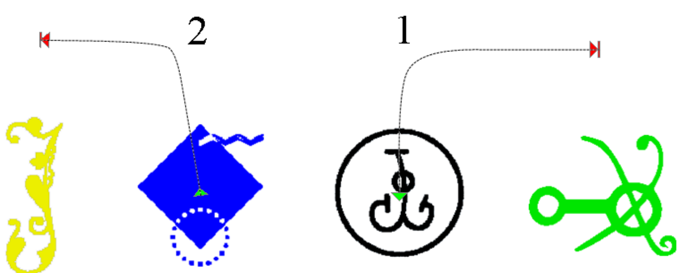


Figure 1. A sample novel-new trial from the experiment. Four objects appeared on the screen at once. After an object moved, the participant would describe the movement.

(1) The chiddle moved above the hamel.
(2) The cammer moved above the neeken.

Procedure

The experiment was presented as a PowerPoint slideshow to the participant on a Macintosh computer. The participant sat in front of the computer in a quiet room, was given a description of the task, and performed two practice trials to ensure compliance with the instructions before beginning the actual experiment. There was no addressee that the participant directed his or her speech to, although the researcher was in the same room.

The experiment was recorded on Praat version 5.3.43 as a mono sound with a sampling frequency of 44,100 Hz (Boersma & Weenink, 2013). The headset microphone was attached to an Audio-Technica Power Module and an Alesis MultiMix 8 FireWire Mixer, with the volume adjusted to the highest setting on the computer. One long recording was made for the entire experiment.

At the completion the experiment, the participant answered a questionnaire that asked about the use of strategies, the creation of the novel words, and the predictability of movements; was debriefed; and was given compensation in the form of course credit or payment. The entire experiment lasted about 45 minutes.

Measures

Target utterances (the second sentence of each trial) were analyzed individually using Praat version 5.3.43 (Boersma & Weenink, 2013). They were excluded if the participant said the wrong noun, if a disfluency (e.g., pauses longer than 250 ms, repairs, fillers like *uh* and *um*) occurred within the target noun, or if the given/new manipulation was no longer maintained due to a disfluency or error in the first sentence. From the waveform, spectrogram, and audio, the following regions were coded and analyzed (Figure 2):

- Onset latency: this is the time from the offset of a beep that co-occurred with each movement to the onset of speech,
- The determiner *the*,
- The target noun, and
    - In addition to the absolute duration, we analyzed this region normalized as a proportion of utterance time, calculated as $\frac{noun}{determiner+noun+verb}$. The normed value controls for speech rate and measures how prominent the target noun is relative to the utterance. This allows for direct comparisons between blocks without having to worry about the raw duration. Even

though all regions are shorter during Block 2, the target noun may be proportionately equivalent to the target noun in Block 1. And even though a difficult target may cause all words in the utterance to be longer, the target may be lengthened proportionately more.

- The verb *moved*.

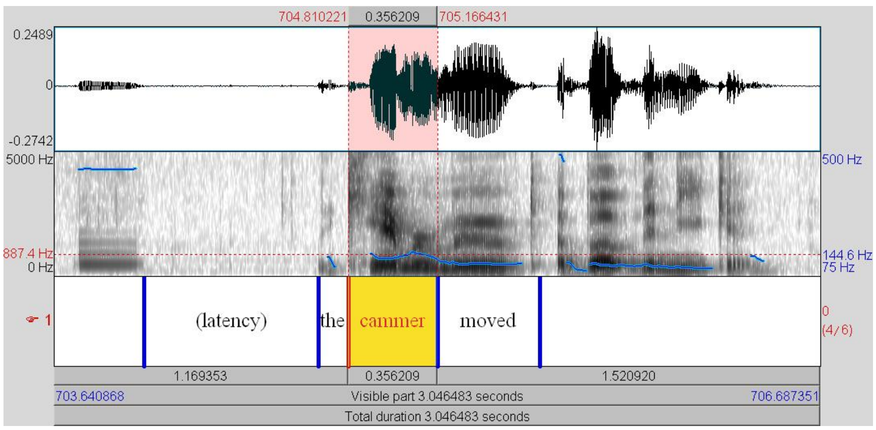


Figure 2. A screenshot of Praat version 5.3.43 during coding. Three raters hand-coded all trials blind to condition and in accordance with the coding criteria in Appendix B.

Interrater Reliability

Three trained individuals in the UNC Language Processing Lab—the first author and two research assistants, RA1 and RA2—coded the aforementioned regions using Praat version 5.3.43 (Boersma & Weenink, 2013). All coding was done blind to condition and followed the criteria in Appendix B. Segmenting stop onsets like [b] is often problematic because there is debate on whether to include the closure time before the stop release. In our coding system, the closure time was not included. This did not affect the relative durations of the conditions because the stimuli were balanced such that a [b] onset appeared in both the familiar and novel wordlists (e.g., *beetle* and *beeler*). There were an equal number of target words with each consonant onset in all conditions.

The first author coded all 24 participants, while the research assistants coded five and three, respectively. The first author's durations were compared to RA1's and RA2's durations. RA1 was not compared to RA2 because they did not code the same participants. Coding was considered reliable if there was less than a 20 ms difference. Extreme differences (greater than 50 ms) were hand-checked and re-coded manually by the first author. Table 1 shows how well the raters matched with each other.

Table 1

Interrater Reliability—The percentages of how often the first author and RA1/RA2 coded the durations similarly (within 20 msec).

Region	First author vs. RA1*	First author vs. RA2**
latency	86%	89%
the	79	84
target	79	79
moved	82	76

Note. *based on five participants
**based on three participants

Results

From a total of 1536 target utterances, 86 were excluded, leaving 1450 for acoustic analysis. Reasons for exclusion include:

- Wrong pronunciation, e.g., *layridge* for *lehridge* (*n* = 25);
- Wrong targets, e.g., *camon* for *cammer* (*n* = 25);
- Manipulation not being maintained, e.g., saying the wrong noun in the first utterance (*n* = 16);
- Missing targets, e.g., using *it* (*n* = 9);
- Repairs, e.g., *caa-cammer* (*n* = 6);
- Wrong verb, e.g., *went over* instead of *moved above* (*n* = 2);

- Pause longer than 250 ms (*n* = 2);
- Unintelligibility (*n* = 1); and

For the latency analysis, an additional 145 trials were excluded because the beep is missing due to technical reasons. For the determiner analysis, an additional four trials were excluded because the participant failed to say *the*. Log-transformed durations were analyzed, although tables and figures show raw durations for ease of interpretation.

The participants saw all trials between the two blocks, so we hoped that would allow for within-subject comparisons. That was not the case because average durations were significantly lower in Block 2. Consequently, two sets of analyses were performed: (a) a three-way ANOVA using the entire experiment with "block" as a third independent variable, and (b) a two-way ANOVA using only Block 1 trials.

Durational Measures

Refer to Table 2 for a summary of average durations for each region per condition, and Figure 3 for graphical displays of onset latency and target duration to see how conditions differed from one another.

Table 2

Means (and Standard Deviations) of the Durations (ms) for Each Region

		Condition			
		Novel		Familiar	
Region	Block	New	Given	New	Given
latency	1	816 (298)	694 (335)	661 (244)	614 (232)
	2	699 (255)	651 (245)	644 (273)	599 (210)
the	1	157 (112)	146 (127)	129 (62)	118 (59)
	2	133 (86)	117 (69)	124 (67)	109 (44)
target	1	478 (165)	413 (126)	408 (102)	383 (86)
	2	408 (104)	393 (102)	394 (104)	379 (89)
moved	1	333 (87)	327 (78)	328 (78)	320 (72)
	2	306 (72)	298 (65)	309 (67)	305 (68)

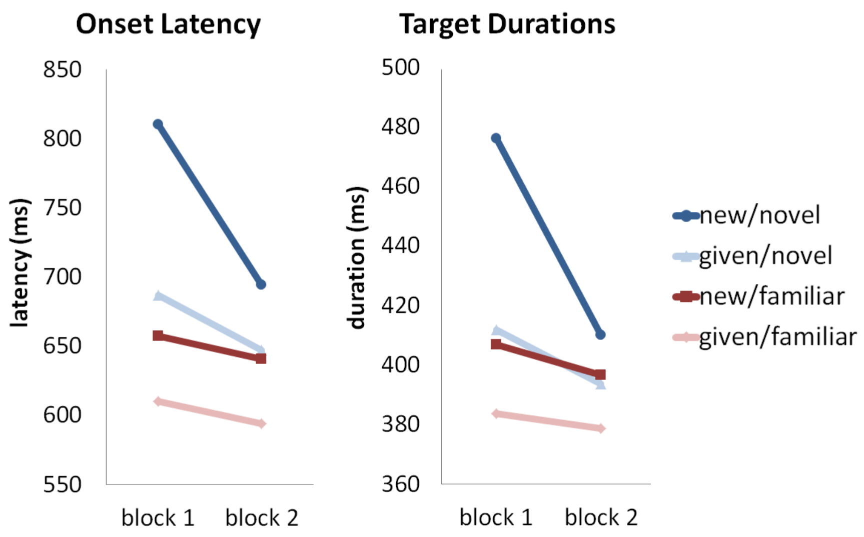


Figure 3. Onset latency and target duration plotted by condition and by block. These graphs show main effects of discourse status, planning difficulty, and block, as well as an interaction effect. Importantly, onset latency pattern with target duration.

Analysis of the entire experiment. A three-way ANOVA (discourse status x planning difficulty x block) was conducted in R. Unless otherwise indicated, the main effects were significant by subjects (*F1*) and by items (*F2*) . Interaction effects are discussed separately for each measure. All *F1*, *F2*, and *p* values are shown in Table 3.

Onset latency. As predicted, there were three main effects: discourse status, planning difficulty, and block. Onset latency was longer before new targets, before novel words, and in Block 1, and shorter before given targets, before familiar words, and in Block 2. All interactions were not significant.

***The determiner* the.** The duration of the determiner patterned similarly to onset latency. This is expected because these two regions together comprise the planning region. It was longer before new targets, before novel words (*F2* was not significant though), and in Block 1, and shorter before given targets, before familiar words, and in Block 2. There was also a significant interaction between planning difficulty and block, such that the longest duration was in the novel trials in Block 1. All other interactions were not significant.

Table 3

Significance for Main and Interaction Effects in the Entire Experiment

Region	Effect	*F1*(1,23)	*p1*	*F2*(1,30)	*p2*
latency	givenness	7.39	.01	12.89	.00
	novelty	11.94	.00	21.82	.00
	block	8.67	.01	7.45	.01
	givenness x novelty	1.41	.25	1.38	.25
	givenness x block	0.45	.51	3.20	.08
	novelty x block	1.83	.19	1.44	.24
	givenness x novelty x block	0.93	.35	0.46	.50
the	givenness	12.50	.00	14.53	.00
	novelty	11.85	.00	1.74	.20
	block	14.87	.00	55.06	.00
	givenness x novelty	1.41	.25	0.74	.40
	givenness x block	0.13	.73	0.28	.60
	novelty x block	10.32	.01	28.06	.00
	givenness x novelty x block	0.28	.60	0.05	.83
target	givenness	21.03	.00	18.75	.00
	novelty	69.16	.00	3.25	.08
	block	7.38	.01	40.82	.00
	givenness x novelty	5.64	.03	1.10	.30
	givenness x block	5.39	.03	1.41	.24
	novelty x block	16.47	.00	14.60	.00
	givenness x novelty x block	3.60	.07	0.81	.38
target/norm	givenness	5.50	.03	8.30	.01
	novelty	5.02	.04	0.28	.60
	block	0.23	.64	1.89	.18
	givenness x novelty	4.12	.06	2.05	.16
	givenness x block	1.91	.18	1.42	.24
	novelty x block	0.39	.54	0.11	.75
	givenness x novelty x block	1.85	.19	2.70	.11
moved	givenness	7.86	.01	1.46	.24
	novelty	0.00	.97	0.07	.80
	block	9.54	.01	69.68	.00
	givenness x novelty	0.93	.35	0.00	.98
	givenness x block	0.01	.92	0.14	.71
	novelty x block	2.30	.15	3.48	.07
	givenness x novelty x block	0.31	.59	0.27	.61

Note. In this table, “givenness” is used to mean discourse status, and “novelty” is used to mean planning difficulty. “target/norm” refers to the target region after it was normalized as a proportion of utterance time. Shaded p values are not significant at $\alpha = .05$.

The target noun. Again, as predicted, there were main effects of discourse status, planning difficulty (*F2* was marginally significant), and block. Given words were shorter than new words; familiar words were shorter than novel words; and Block 1 words were shorter than Block 2 words. Discourse status interacted with planning difficulty (*F2* was not significant though) such that the longest target duration was for new and novel targets. As shown in Table 4, discourse status had a greater effect on novel targets (~15%) than familiar targets (~7%), as quantified by percent difference, but only in Block 1. In other words, difficult words are reduced more than easy words. Additionally, the robust effect of block was manifested in interactions between discourse status and block (*F2* was not significant though), as well as between planning difficulty and block.

Table 4

Effects of Discourse Status on Novel and Familiar Targets, as Quantified by Percent Difference

Condition		Block 1		Block 2	
		Difference in onset latency	Difference in target duration	Difference in onset latency	Difference in target duration
Novel	New Given	16.5%	14.5%	7.0%	4.1%
Familiar	New Given	7.4%	5.9%	7.6%	4.6%

Note. To read this table: The onset latency in the novel-given condition is 16.5% shorter than the onset latency in the novel-new condition.

The target noun as a proportion of utterance time ($\frac{noun}{determiner+noun+verb}$). When assessing the relative prominence of the target noun, there were still main effects of discourse status and planning difficulty. However, block no longer had a main effect, and the interactions involving block were no longer significant either.

***The verb* moved.** There were main effects of discourse status (*F2* was not significant though) and block, but not of planning difficulty. This was expected because (a) *moved* is a repeated word, (b) participants likely went faster during Block 2, and (c) the difficulty of planning the target should not influence the production of the verb, which followed the target and was not in its planning region. All interactions were not significant.

Analysis of Block 1 only. A two-way ANOVA (discourse status x planning difficulty) was also conducted in R, where block/practice effects were not taken into consideration. Overall, this analysis mirrored that of the entire experiment, so it will not be further discussed. All *F1*, *F2*, and *p* values are shown in Table 5.

Table 5

Significance for Main and Interaction Effects in Block 1 Only

Region	Effect	*F1*(1,23)	*p1*	*F2*(1,30)	*p2*
latency	givenness	6.68	.02	10.84	.00
	novelty	10.39	.01	21.98	.00
	givenness x novelty	3.33	.09	1.27	.27
the	givenness	0.74	.40	4.85	.04
	novelty	18.22	.00	4.68	.04
	givenness x novelty	0.80	.38	0.08	.78
target	givenness	17.13	.00	9.24	.01
	novelty	56.72	.00	6.35	.02
	givenness x novelty	5.18	.04	1.23	.28
moved	givenness	3.13	.09	0.40	.53
	novelty	0.60	.45	1.49	.23
	givenness x novelty	0.06	.81	0.06	.81

Note. In this table, "givenness" is used to mean discourse status, and "novelty" is used to mean planning difficulty. Shaded p values are not significant at α = 05.

Discussion

These results showed clear main effects of discourse status and planning difficulty on acoustic reduction. When the target was new (i.e., it has not been evoked yet) or novel (i.e., it is difficult to plan), the onset latency, determiner, and target noun were longer than when the target was given or familiar. These main effects replicated findings from past research, so our paradigm here is sensitive enough to pick up these measures. However, it is important to note that some of the by-items analyses did not reach statistical significance; this problem is discussed again later in the section.

The variation in latency and determiner, which together comprise the planning region, can be explained by accounts of incremental planning. They propose that speakers lengthen this region in order to give themselves more time to retrieve and produce difficult words (Bell et al., 2009). The results here follow that reasoning. New targets had not been evoked in the discourse, and novel words were abstract, so they were difficult for the speaker. As such, these target nouns were preceded by a longer latency and determiner. These latency effects for given targets support the idea that givenness is related to planning facilitation. Assuming only an effect of discourse status, then the target noun would get shortened, but nothing else. However, the fact that the planning region is shortened suggests that a given referent is also easier to retrieve and produce. This allows the speaker to devote more resources to forming the utterance, thus resulting in faster articulation.

The variation in target duration can be attributed to two accounts: (a) the pragmatic selection rule, or (b) speaker facilitation. The pragmatic selection rule maintains that given referents are reduced/shorter, while new referents are prominent/longer. This exact pattern was shown by the main effect of discourse status. The second account, speaker facilitation, states that words that are easy to retrieve and produce (e.g., familiar words) tend to be reduced/shorter, while harder words tend to be prominent/longer. This pattern was shown by the main effect of planning difficulty.

Everything mentioned until now was expected. More interesting is the interaction between discourse status and planning difficulty, which gets at the critical question of how familiar and novel words are affected by repeated mention. The first mention of a novel word is predicted to have the longest onset latency and target duration because of a greater cognitive load required to identify the abstract drawing and recall the appropriate name. We further hypothesized that this extra difficulty in planning should not have as great of an effect during the repeated mention—the representation is already accessed and stored in working memory, in the same place where a familiar word would be stored. Therefore, it was expected that given novel words would have similar durations as given familiar words.

That was not what we observed. The novel words were reduced to a greater degree than familiar words (Table 4), but they were never reduced to the same level. Instead, we saw that the effect of planning difficulty (novelty) persisted: Even though the first- and second-mention of the novel word occurred within a few seconds of each other, there was still difficulty in producing the repeated novel word. Given novel words were shorter than new novel words (main effect of discourse status), but they were still longer than given familiar words (main effect of planning difficulty).

So our original hypothesis was partially supported by these data: There was a trend towards an interaction between discourse status and planning difficulty—that repeated mention affected novel words greater than it did familiar words—but not to the degree where repeated novel words would have the same durations as repeated familiar word. Nevertheless, the interaction suggests that discourse status and planning difficulty work together to affect acoustic reduction, and that the former is mediated by the latter. If the two factors were indeed independent, then we would not have gotten the asymmetrical reduction between familiar and novel words, and instead, they may have been reduced by the same proportion (as opposed to absolute duration). That is, reduction due to repeated mention would not be contingent on the word's planning difficulty.

Another interesting result is that onset latency patterned in the same way as target duration (Figure 3), even though the interaction was significant only for target duration and only by subjects. In other words, onset latency, a measure of facilitation/planning difficulty, is correlated with target duration, a measure of discourse status. This seems to provide further support that discourse status and planning difficulty are not independent (as suggested by Arnold & Watson, under review).

Furthermore, there was a significant effect of block, such that Block 2 durations were consistently shorter for all conditions. This means that the participants got faster over the course of the experiment, which was expected because (a) they became accustomed to the task and (b) the novel words became less "novel" due to practice effects. Despite this general speed-up, the main effects of discourse status and planning difficulty were still evident, such that given targets and familiar words were shortened comparatively more than the rest of the utterance. In addition, discourse status and planning difficulty both interacted with block, but only at the determiner and target noun regions. These interactions suggest that the difference between given/new and novel/familiar gets smaller in Block 2; however, this might be a floor effect (i.e., the word cannot be shortened anymore that it already has).

When speech rate was controlled, the effects of discourse status and planning difficulty remained. Because we only manipulated the discourse status of the target, givenness was expected to affect the target region only. Therefore, it should be shortened in the given condition, regardless of what the speech rate was. The novelty manipulation, on the other hand, could affect speech rate in general and slow down the entire utterance, which would still explain why novel words were longer than familiar words. However, the fact that planning difficulty was still significant (albeit marginally) even when speech rate was controlled suggests that novel targets were slowed proportionally more than the rest of the utterance.

This experiment overall presented with somewhat favorable results that replicated known findings, showed that novel words are reduced by a larger proportion than familiar words, and suggested that discourse status is partially mediated by planning difficulty. Unfortunately, the by-items (*F2*) analysis failed to reach significance for some of the measures, which was likely due to the between-items design. If the lack of significance were not just an effect of power, then that would mean the experimental items cannot generalize to all English words and all nonwords that sound like English words. Alternatively, and more worrisome, the lack of significance might suggest that the observed effects are driven by a subset of our items.

To get a better understanding of these data, pitch and intensity analyses need to be conducted (see Arnold & Tanenhaus, 2012; Christodoulou, 2009; Isaacs & Watson, 2010, for a discussion on pitch and prosody results). These measures may be more informative, especially because they are not affected by speech rate like duration is. Duration is only one indicator of acoustic prominence, so pitch and intensity may confirm the current results or even tell a different story. With this experiment, we might expect to see greater pitch movement and higher intensities in the new/novel referents, and flatter contours and lower intensities for given/familiar referents.

And lastly, to attain a better understanding of the interaction between discourse status and planning difficulty, further research needs to be done. Different factors can be manipulated, such as predictability or compounding. In progress now is an experiment identical in design that looks at in-focus (i.e., the target is the subject of both utterances) and focus-shift (i.e., the target is the object in the first utterance, but becomes the subject in the next one) trials, as in (5) and (6), respectively.

(5) The cammer moved above the hamel.

The <u>cammer</u> moved above the neeken.

(6) The chiddle moved above the cammer.

The <u>cammer</u> moved above the neeken.

Terken and Hirschberg (1994) claim that reduction is even greater when the repeated mention word has the same grammatical role as the first mention. In this example, that means the second *cammer* in (5) will be more reduced than the second one in (6). This is predicted because although the *cammer* is given in both trials, it will be made slightly more prominent in (6) to emphasize a contrast in focus. We know this to be true with real words, but with the novel stimuli, we can once again ask how planning difficulty interacts with focus to affect acoustic reduction. Only by analyzing the data here will we find out.

References

Arnold, J.E. & Tanenhaus, M.K. (2012). Disfluency isn't just um and uh: The role of prosody in the comprehension of disfluency. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Arnold, J.E. & Watson, D.G. (under review). Synthesizing meaning and processing approaches to prosody: Performance matters. Ms., University of North Carolina at Chapel Hill.

Balota, D.A., Boland, J., & Shields, L. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, *28*(1), 14-36. doi:10.1016/0749-596X(89)90026-0.

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

Bard, E.G., Lowe, A., & Altmann, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), *Proceedings of EUROSPEECH Vol. 2* (pp. 573-576). Paris, France: European Speech Communication Association.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636-1657. doi:10.1016/j.pragma.2005.03.017.

Baumann, S., & Hadelich, K. (2003). Accent type and givenness: an experiment with auditory and visual priming. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1811–1814). Barcelona, Spain.

Bell, A., Brenier, J., 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*(1), 92–111. doi:10.1016/j.jml.2008.06.003.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.43, retrieved from http://www.praat.org/.

Brown, G. (1983). Prosodic structures and the given/new distinction. In D.R. Ladd & A. Cutler (Eds.), *Prosody: Models and measurements* (pp. 67–77). Berlin, Germany: Springer.

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Christodoulou, A. (2009). *Thinking prosody: How speakers indicate production difficulty through prosody*. (Unpublished master's thesis), UNC Chapel Hill.

Christodoulou, A. (2012). *Variation in word duration and planning*. (Unpublished doctoral dissertation), UNC Chapel Hill.

Davies, M. (2008-). The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57-84. doi:10.1006/jmla.2001.2797.

Forster, K.I., & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6): 627–635. doi:10.1016/S0022-5371(73)80042-8.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, *31*(4), 307-319. doi:10.1177/002383098803100401.

Fowler, C., & 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*(5), 489-504. doi:10.1016/0749-596X(87)90136-7.

Griffin, Z.M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10*(3), 603-609. doi:10.3758/BF03196521.

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

Halliday, M.A.K. (1967). Notes on transitivity and theme in English, *Journal of Linguistics*, *3*(2), 199-244. doi:10.1017/S0022226700016613.

Hayes, B. (2012). BLICK: a phonotactic probability calculator [Computer program]. Version 1.0, retrieved from http://www.linguistics.ucla.edu/people/hayes/BLICK/.

Isaacs, A., & Watson, D.G. (2010). Accent detection is a slippery slope: Direction and rate of F0 change drives listeners' comprehension. *Language and Cognitive Processes*, *25*(7-9), 1178-1200. doi:10.1080/01690961003783699.

Jurafsky, D, Bell, A., Gregory, M., and Raymond, W.D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam, Netherlands: John Benjamins.

Kahn, J. & Arnold, J.E. (2012) A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*(3), 311-325. doi:10.1016/j.jml.2012.07.002.

Kahn, J. & Arnold, J.E. (under review). Speaker-internal processes drive durational reduction. Ms., University of North Carolina at Chapel Hill.

Meyer, A, Belke, E., Hacker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, *57*(2), 210-231. doi:10.1016/j.jml.2006.10.005.

Meyer, A.S., Roelofs, A., & Levelt, W.J.M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*(1), 131-147. doi:10.1016/S0749-596X(02)00509-0.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118, 2561–2569. doi:10.1121/1.2011150.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart’s object database: Color and texture improve object recognition. *Journal of Vision*, *1*(3), 413. doi: 10.1167/1.3.413.

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*(2), 174-215. doi: 10.1037/0278-7393.6.2.174.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing ‘‘given’’ information: Effects of persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125-145. doi:10.1177/002383099403700202.

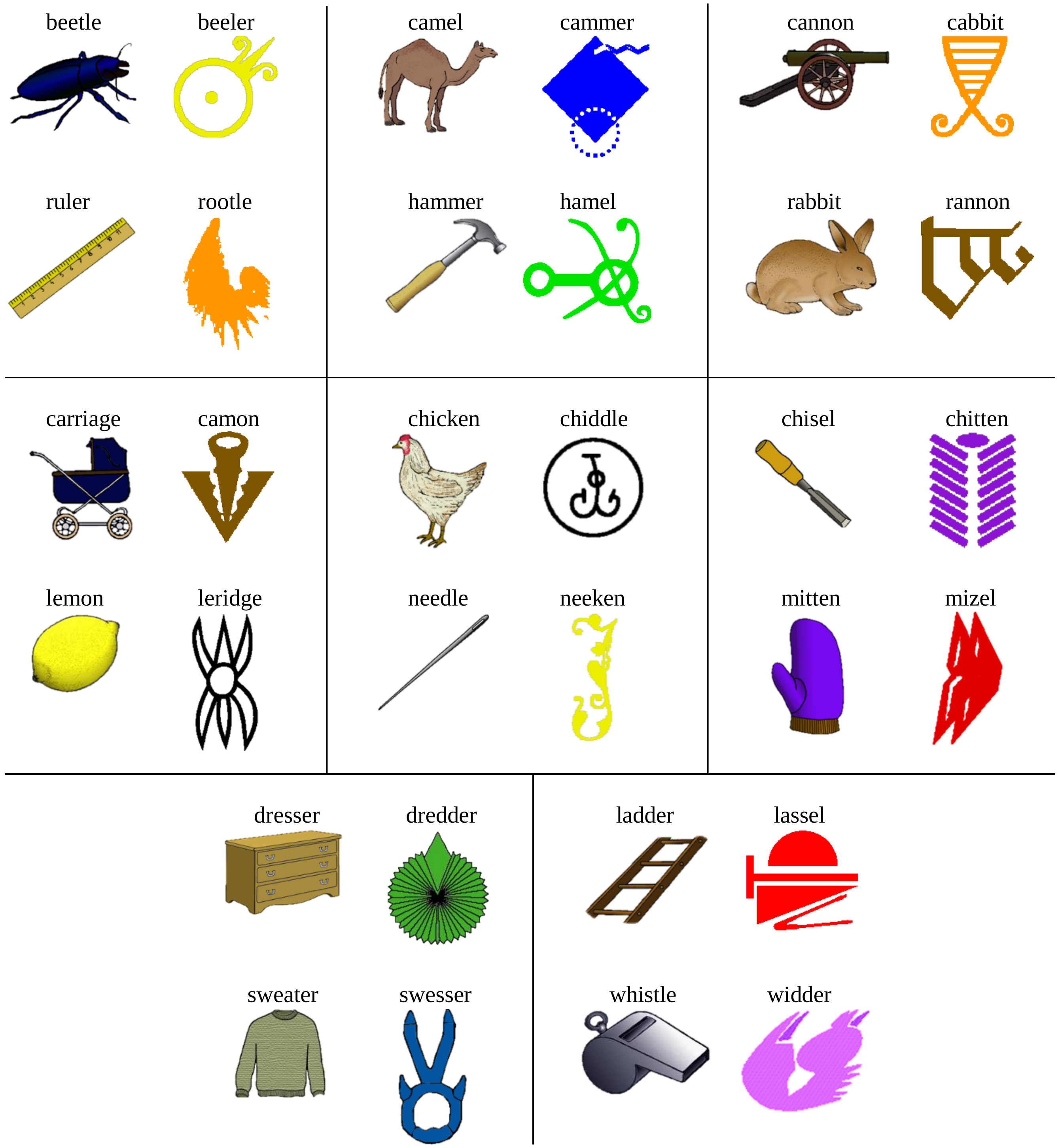
Vitevitch, M.S. & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, *36*(3), 481-487.

Watson, D.G. (2010). The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, *52*, 163 -183. doi:10.1016/S0079-7421(10)52004-8.

Zipf, G.K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Boston, MA: Houghton Mifflin Company.

Appendix A
Stimuli

Each group of objects consists of two familiar words and the two novel words that result from their recombination.



Appendix B
Coding Criteria

When coding each sound file, four regions were demarcated: (a) the onset latency, (b) the determiner *the*, (c) the target noun, and (d) the verb *moved*. Furthermore, depending on when it occurs, a disfluency may be put into a region of its own. The criteria used in identifying these regions are as follows. The *waveform* refers to the top line, and the *spectrogram* refers to the second line.

(a) The onset latency {SL} . . .

is the region after the target begins moving on the computer screen (as indicated by a beep {NS}) and before the participant begins speaking, as shown in Figure B1. A disfluency such as *uh* and *um* that occurred in this region is kept as part of this region.

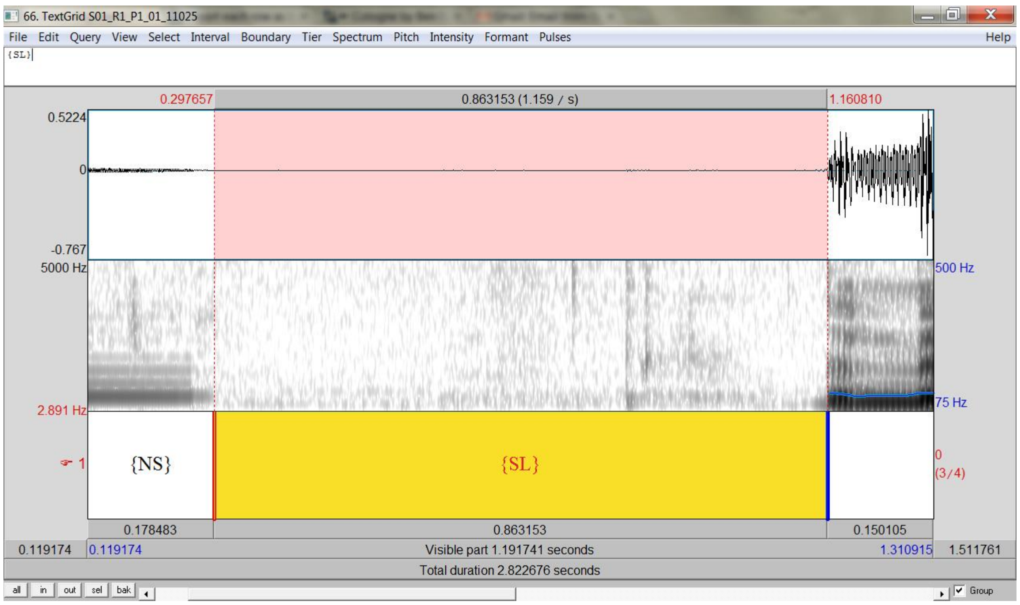


Figure B1. The onset latency's left edge is the end of the beep, and the right edge is the beginning of *the*.

(b) The determiner *the* . . .

is marked by a noticeable difference in the waveform and spectrogram that corresponds with hearing the word *the* (i.e., not background noise or disfluencies). Pre-voicing, which is characterized by an extremely low amplitude waveform and glottal pulses at the bottom of the spectrogram, is not included in this region. Figures B2 and B3 show the presence and absence of pre-voicing, respectively. The end of *the* is typically the beginning of the target noun, which is described in (c). If there is a disfluency (like *uh*, *um*, and/or a pause greater than 250 ms) after *the* and before the noun, then it will receive its own region labeled as {SP1}.

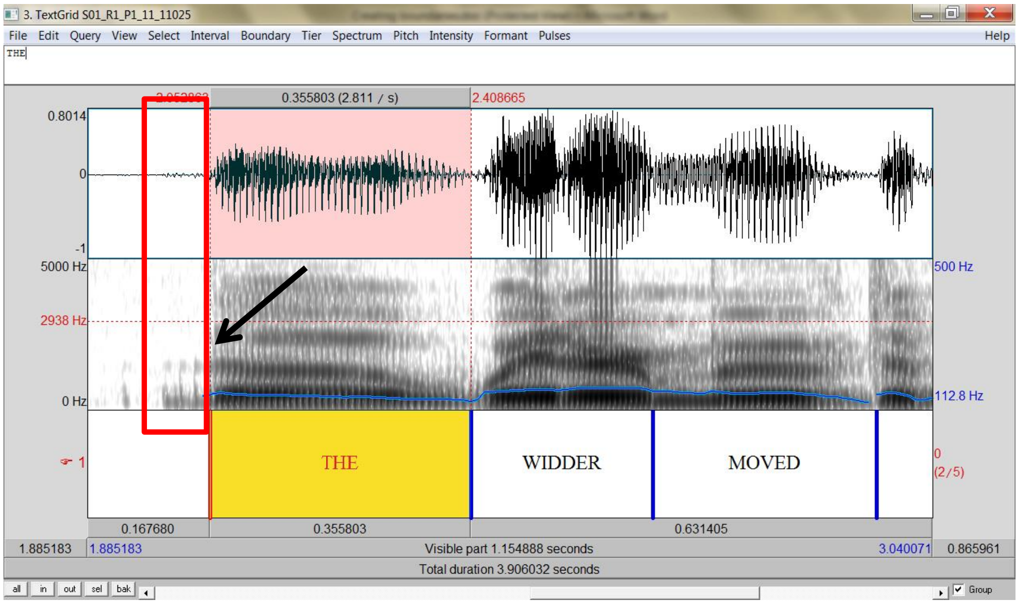


Figure B2. The left edge of *the* is typically marked by a sudden darkening in the spectrogram, as shown by the black arrow. Pre-voicing, which is enclosed in the red box, is not included as part of the determiner region.

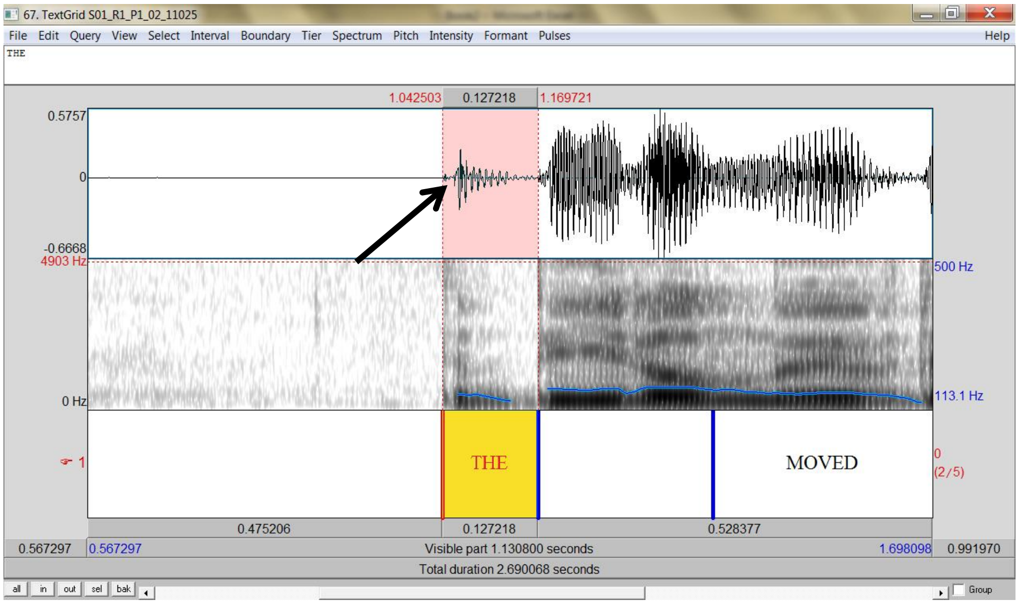


Figure B3. The waveform alone is not sufficient when coding. This determiner region here begins with a low amplitude waveform, as shown by the black arrow. Although this is a characteristic of pre-voicing, the spectrogram is noticeably darker, which sugges ts that this is indeed the onset of *the*.

(c) The target noun . . .

is the most variable region because there are 32 possible words. Depending on the first consonant, there are different landmarks that need to be considered. The end is easier to detect because it is typically the beginning of the verb *moved*, which is described in (d). If there is a disfluency after the noun and before *moved*, then it will receive its own region labeled as {SP2}. If the disfluency occurs in the middle of the noun, then the entire trial is excluded.

For *beetle*, *beeler*, *camel*, *cammer*, *cannon*, *cabbit*, *carriage*, and *camon*, the stops [b] and [k] begin at the stop release (i.e., when sound is finally heard). The waveform is aperiodic, and the spectrogram goes from light to dark at this point, as shown by the black arrow in Figure B4.

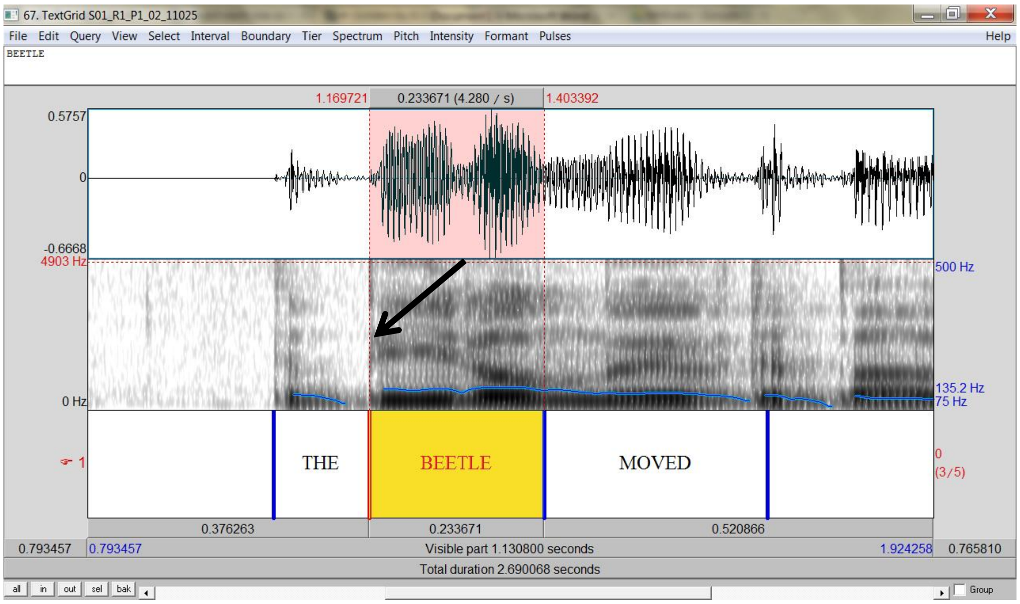


Figure B4. The left edge of the target nouns that begin with [b] and [k] begins at the stop release, as indicated by the black arrow, and the right edge is the beginning of *moved*.

For *chicken*, *chitten*, *chisel*, *chiddle*, *dresser*, *dredder*, *sweater*, and *swesser*, the affricates [ʧ] and [ʤ] and fricative [s] begin when spectrogram goes from light to dark at this point, particularly at high frequencies, as shown by the black arrow in Figure B5 and B6. Additionally, there should be no glottal pulses at the bottom of the spectrogram.

Furthermore, for *chicken*, *chitten*, *cannon*, *camon*, *lemon*, *mitten*, *neeken*, and *rannon*, unless there are clear formant changes that mark the end of [n], the offset of the target is the halfway point between that [n] and the [m] from the following *moved*.

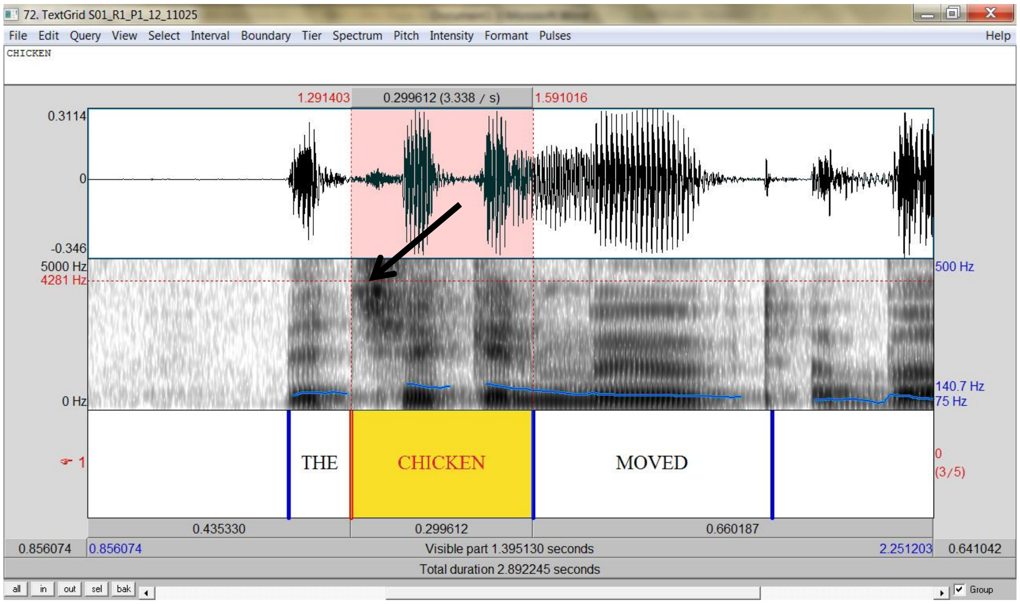


Figure B5. The left edge of the target nouns that begin with [ʧ] and [ʤ] begins at the dark region of high frequencies, as indicated by the black arrow, and the right edge is the beginning of *moved*.

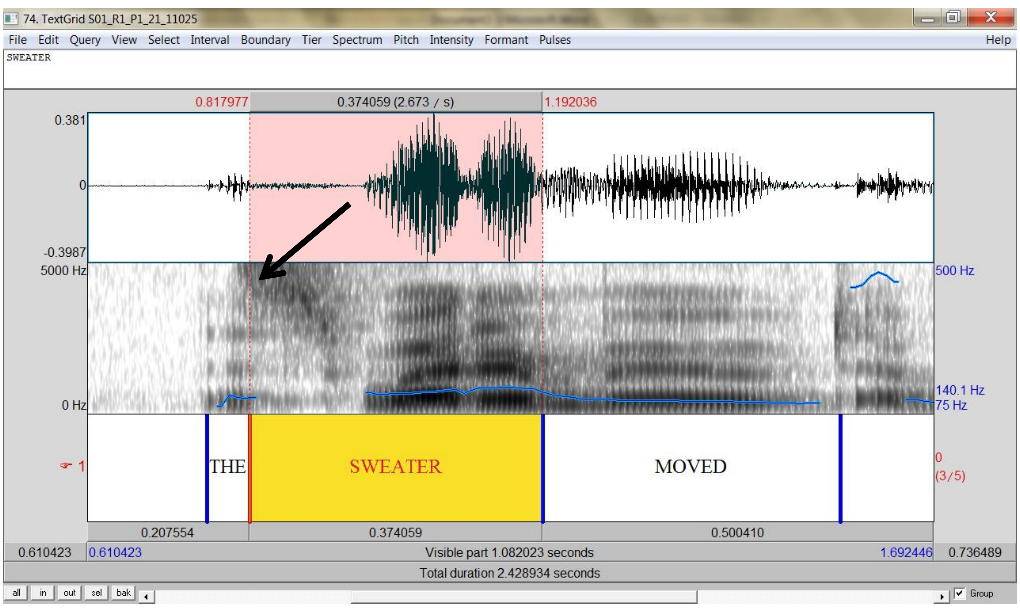


Figure B6. The left edge of the target nouns that begin with [s] begins at the dark region of high frequencies, as indicated by the black arrow, and the right edge is the beginning of *moved*.

For *mitten*, *mizel*, *needle*, and *neeken*, the nasals [m] and [n] the waveform is somewhat lower amplitude and the spectrogram is slightly fainter compared to the surrounding sounds, as indicated by the black arrows in Figure B7.

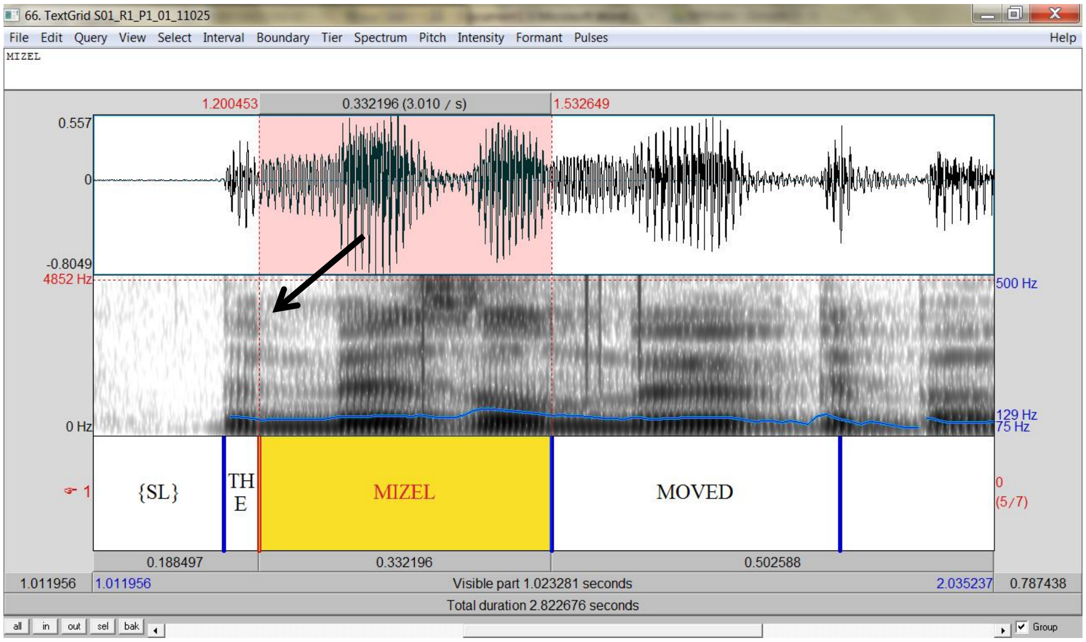


Figure B7. The left edge of the target nouns that begin with [m] and [n] is low amplitude (in the waveform) and faint (in the spectrogram), as shown by the black arrows. The right edge is the beginning of *moved*.

For *ladder*, *lassel*, *lemon*, *leridge*, *rabbit*, *rannon*, *ruler*, *rootle*, *widder* and *whistle*, the approximants [l], [r], and [w] are extremely tricky. These sounds begin when the formants (the contours on the spectrograms) reach its lowest steady state, as indicated by the red bar in Figures B8.

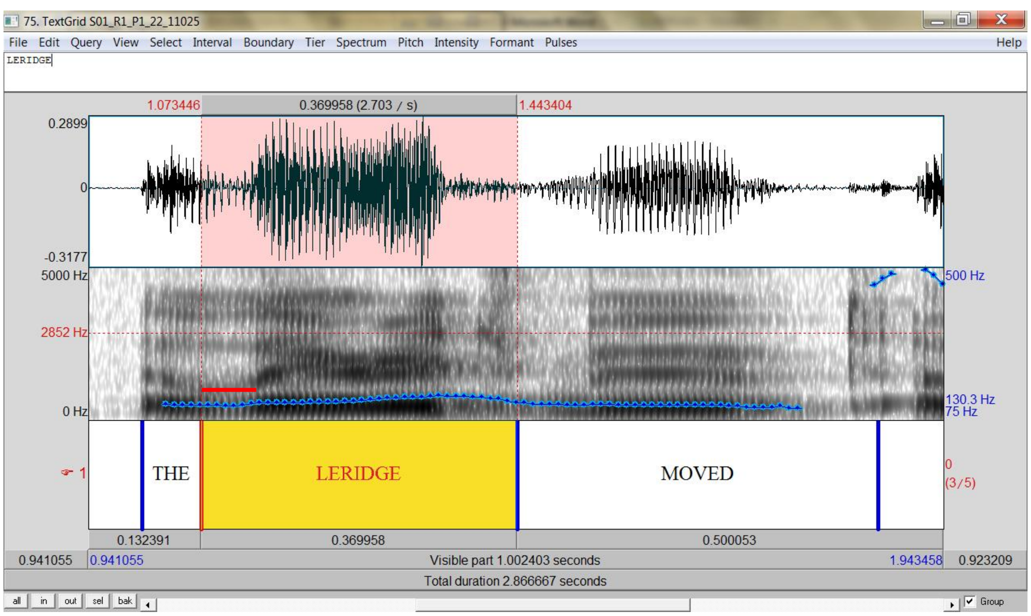


Figure B8. The left edge of the target nouns that begin with [l], [r], and [w] begins when the formants (the contours on the spectrograms) reach its lowest steady state, as indicated by the red bar. The right edge is the beginning of *moved*, as always.

(d) The verb *moved* . . .

begins when the waveform is somewhat lower amplitude and the spectrogram is slightly fainter compared to the surrounding sounds, as indicated by the black arrows in Figure B9. This region ends with a short stop burst, as shown in the red box. Note that this stop burst may or may not be present, so it is a good idea to listen for the onset of the next word *above*. A disfluency after the verb is ignored.

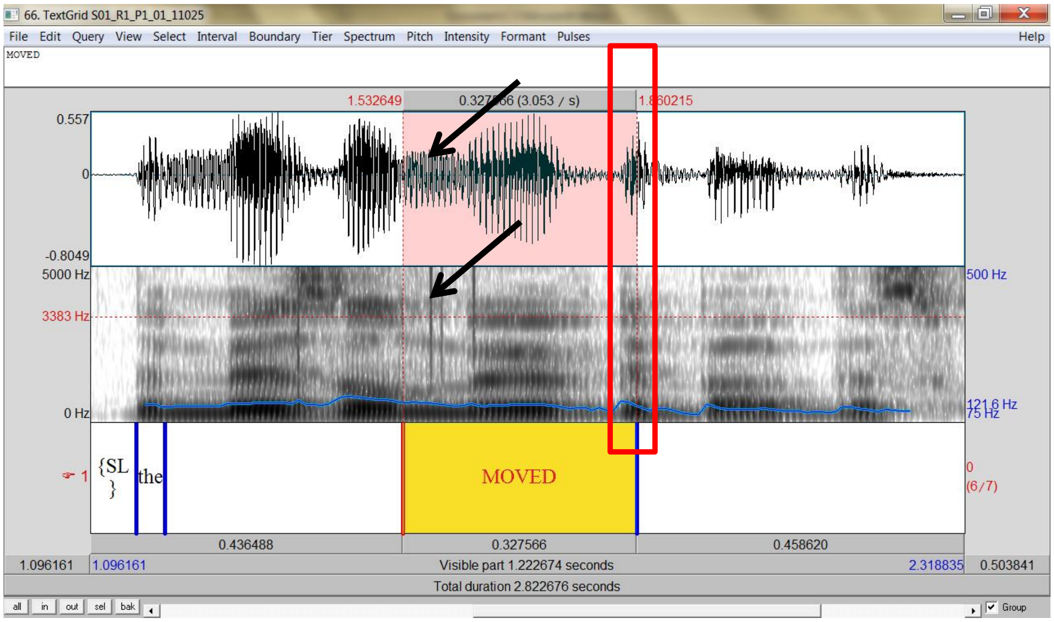


Figure B9. The verb region's left edge is low amplitude (in the waveform) and faint (in the spectrogram), as shown by the black arrows. The right edge is a short stop burst, as shown by the red box.