UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SPEAKING WHILE THINKING: INCREMENTAL ARTICULATION IN MEMORY RECALL TASKS

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Bryan B. Holbrook

December 2017

The Dissertation of Bryan B. Holbrook is approved:

______________________________
Professor Alan H. Kawamoto, chair

______________________________
Professor Jean E. Fox Tree

______________________________
Professor Benjamin C. Storm

Tyrus Miller
Vice Provost and Dean of Graduate Studies
# Table of Contents

List of Tables .................................................................................................................. v  
List of Figures ................................................................................................................... vi  
Abstract ............................................................................................................................ vii  
Acknowledgments ............................................................................................................. viii  

Introduction ...................................................................................................................... 1  
Incremental Articulation .................................................................................................... 2  
Relation Between Speech Production and Memory Recall .............................................. 6  
  Phonological Similarity Effect ....................................................................................... 6  
  Planning Scope .............................................................................................................. 7  
  Chunking ....................................................................................................................... 8  
  Temporal Grouping ....................................................................................................... 9  
Chronometric Measures in Recall Tasks ........................................................................ 10  
Summary ........................................................................................................................... 14  
Experiment 1 .................................................................................................................... 15  
  Method ........................................................................................................................... 17  
    Participants ................................................................................................................. 17  
    Materials ...................................................................................................................... 17  
    Apparatus ................................................................................................................... 19  
    Procedure .................................................................................................................... 19  
  Results ........................................................................................................................... 20  
  Discussion ..................................................................................................................... 25  
Experiment 2 .................................................................................................................... 28  
  Method ........................................................................................................................... 29  
    Participants ................................................................................................................. 29  
    Materials and Apparatus ............................................................................................ 30  

iii
List of Tables

Table 1. Means and standard deviations in milliseconds for ISD_Remark and as a proportional difference for ISD_prop_diff in Experiment 1................................. 24

Table 2. Final model for ISD_prop_diff in Experiment 1. Bolded parameters are significant at $p < .05$. Parameter estimates are in seconds. $N_{\text{participants}}=14$. $N_{\text{items}}=50$.................................................................................................................. 24

Table 3. Means and standard deviations for ISD_prop for the relevant independent variables in Experiment 2 ................................................................. 32

Table 4. Final models for ISD_prop in Experiment 2. Bolded parameters are significant at $p < .05$. Parameter estimates are in seconds. $N_{\text{participants}}=60$; $N_{\text{items}}=50$. ..... 34

Table 5. Means and standard deviations of recall scores and error rates in Experiment 2......................................................................................................... 37

Table 6. Means and standard deviations in milliseconds for recall times (ISD and RWD) and as a proportional difference for ISD_prop_diff for the relevant independent variables in the additional analyses for Experiment 2.................. 56

Table 7. Final models for ISD and RWD during recall, and ISD_prop_diff in additional analyses for Experiment 2. Bolded parameters are significant at $p < .05$. Parameter estimates are in seconds. $N_{\text{participants}}=60$; $N_{\text{items}}=50$. ...................... 57
List of Figures

Figure 1. Boxplot for ISD during recall for both conditions in Experiment 1.......... 22
Figure 2. Boxplot for ISD_{prop,diff} for both conditions in Experiment 1.................... 23
Figure 3. Boxplot for ISD_{prop} for both conditions in Experiment 2......................... 31
Figure 4. Key comparison in determining overall effect of between-block interference in Experiment 2. Error bars represent one standard error of the mean.......... 38
Abstract

Speaking While Thinking: Incremental Articulation in Memory Recall Tasks

by

Bryan B. Holbrook

Examining verbal memory recall from a speech production standpoint can greatly contribute to the understanding of both processes. The existence of the segment as the minimal unit of articulation, already shown in the naming task, has important implications for how recall might occur in verbal recall tasks. Specifically, the initial segment of a target response might be retrieved and articulated before the remainder of the word is retrieved, leading to incremental articulation of the target word with the segment as minimal unit of articulation. This possibility was explored in two experiments by examining the initial segment durations of responses during recall when blocks were either heterogeneous or homogeneous with respect to the initial segment. Incremental articulation based on the segment was found to occur during the recall task to a greater degree in homogeneous blocks in both experiments. Experiment 2 showed no strong evidence for a relation between interference and incremental articulation. Additional evidence for the segment as minimal unit of articulation was found, extending this result beyond the naming task. Prior techniques used to explore the time-course of recall, such as interresponse time, must be revised. Through exploring recall using this novel paradigm inspired by speech production research, new methods can be utilized to examine established recall effects such as the phonological similarity effect.
Acknowledgments

First and foremost, I would like to thank my advisor Dr. Alan H. Kawamoto, for his mentorship and help developing my academic skills over the past years. Thanks so much for believing in me and always being available to chat. I would also like to thank my dissertation committee, Dr. Jean E. Fox Tree and Dr. Benjamin C. Storm for their feedback and thoughts on this dissertation and helping me throughout its development, my qualifying exam, and graduate school in general. I would also like to thank Dr. Qiang Liu for his work in our lab prior to my arrival, specifically for his work designing and implementing our data capture technique and writing our very useful Praat algorithms. Thanks also to Peter Krause for discussing theoretical and methodological issues with me, and being a great lab partner and friend – I’m sure I’ll be reading your dissertation soon! Thanks to all my research assistants; in particular, Karen Annell, Kellyann Kelso, Maria Russo, Jeremy Santiago, Brisa Speier-Brito, and Nancy Tieu for their help with data collection and analysis on this project. Big thanks to the Psychology administration, especially Allison Land for her guidance and help with acquiring teaching opportunities, ensuring I could continue to attend graduate school. Lastly, I want to extend my heartfelt thanks to my wife Jamie, my mom Libby, my dad Steve, and all my fellow psychology grads for their unwavering support, friendship, and advice throughout the years: thanks for having more faith in me than I had in myself at times – you were right! 
Speaking While Thinking: Incremental Articulation in Memory Recall Tasks

Although very little conscious effort is put into articulation when compared to other cognitive tasks, the incredible complexity of speech production is fully appreciated when attempting to explain the multitude of processes involved. When speech production is operating smoothly, it can be very easy to forget just how complicated a task it is. While progress has been made in understanding how humans perceive and produce language, these abilities are not yet fully explained.

The use of language allows for many cognitive operations, including displaced learning through reading, distributed cognition through writing information down, and symbolic processing such as mathematics and abstract reasoning. Language is also tied to advanced abilities in remembering and recalling information. While clearly not all functioning of memory is dependent on language, the use of language is a key component of semantic knowledge, short-term recall, and verbal working memory. Even when performing tasks which do not strictly require verbal responses, such as remembering what groceries you need to pick up from the store, covert rehearsal and covert articulation are likely occurring. This covert speech is inherently tied to the behavioral processes involved with overt speech production. In order to best understand all cognition, including memory, an understanding of language, particularly speech production, is necessary.

The present work focuses on developing a novel empirical methodology for examining issues in both speech production and memory recall. New theories in
speech production can be applied to tasks and processes not typically considered to involve articulation. While the two fields are clearly linked – at least when considering verbal recall – many research programs and methodologies tend to isolate from one another. Research in both fields can be stimulated by applying current theories of speech production to memory recall.

**Incremental Articulation**

Incremental articulation is the ability of a speaker to prepare upcoming articulatory information while speech production is already occurring. This ability allows for people to begin to speak an utterance before every word or sound in that utterance has been fully planned. For example, an improvised speech can be incrementally articulated sentence by sentence or phrase by phrase (or even word by word, for those who like to live dangerously) – the whole speech need not be planned out meticulously before articulation begins. This can then result in articulatory variation during an utterance due to processing difficulty, such as incomplete target information (Kawamoto, Liu, Lee, & Grebe, 2014), upcoming numerical calculations (Ferreira & Swets, 2002), or interference in lexical access (Fink, Oppenheim, & Goldrick, 2018).

In Leveit’s model of speech production (Leveit, 1989; Leveit, Roelofs, & Meyer, 1999), information is passed from stage to stage incrementally; that is, information is fed forward as it becomes available, and does not wait for previous stages to terminate to proceed to the next stage. The smallest item which must be fully encoded before being passed on, or the minimal unit of processing, varies
between stages. Theories of the minimal unit between phonological encoding and articulation (the minimal unit of articulation) have varied throughout the years, from the level of the phonological word (Levelt et al., 1999) to the syllable (Schriefers & Teruel, 1999) to the segment (Kawamoto, Kello, Jones, & Bame, 1998). Currently, the syllable is thought to be the minimal unit of articulation in Levelt’s model (Meyer, Roelofs, & Levelt, 2003). Thus, a syllable must be fully phonologically encoded before being sent downstream – articulation does not accept single segments as input. The major empirical support for such claims comes from studies of response latency in various naming tasks (Meyer, 1990; 1991; Roelofs, 2002; 2004).

The logic of using latency as a measure of processing time is straightforward. If a certain task takes longer to do than another comparable task, then the slower task involves more resources or more interference resulting in more processing difficulty than the faster one (Sternberg, 1969; Meyer, Osman, Irwin, & Yantis, 1988). However, certain issues arise when using acoustic latency as a measure of response latency which complicate the matter. While commonly determined by a voice key, acoustic onset can also be determined through digital analysis of the full recorded audio waveform itself. This more precise and less phonetically biased method of determining acoustic latency is necessary due to voice key detection issues within and between classes of phonemes (Pechmann, Reetz, & Zerbst, 1989; Kessler, Treiman, & Mullennix, 2002). The transition to offline recorded analysis has another benefit; namely, it allows for additional quantitative and qualitative information about the verbal response to be gathered.
Duration measurements can be useful in providing information about ongoing processing occurring concurrently with action in any task; longer durations indicate delays due to upcoming difficulty in planning or encoding the remainder of the response, specifically the presence of incremental articulation. Acoustic duration has been explored throughout psycholinguistic and working memory research with a focus typically on a large unit size, such as sentence duration (e.g., Ferreira & Swets, 2002) or duration of multisyllabic words (e.g., Sternberg, Monsell, Knoll, & Wright, 1978). However, segment durations can also be useful in exploring the time-course of production (Kawamoto et al., 1998). Evidence has been found supporting duration variability at the level of the segment based on experimental conditions (Kawamoto et al., 2014). Given this, segment duration can be used as a valuable tool in exploring verbal responses as elongated segment durations are taken as an indicator of difficulty in planning or encoding subsequent response information.

After accounting for the phonetic confounds involved with acoustic latency and by measuring segment duration as well as latency, mounting evidence has been found that supports the segment as the minimal unit of articulation (Kawamoto, Liu, & Kello, 2015). For example, Kawamoto et al. (2014) found shorter acoustic latencies and longer initial segment durations when priming the initial segment in a pairwise prime naming task. In addition, non-error trials were found with negative acoustic latencies when the initial segment was primed, indicating acoustic production began before the entire word was presented on a subset of trials. These productions were not ubiquitous however – such trials were found amongst certain participants and
completely absent amongst others. This split between participants was likely due to the adoption of differing response criteria, with some participants adopting a segment criterion (begin articulation once the initial segment is phonetically encoded) and others adopting a word criterion (begin articulation only once the entire word is phonetically encoded), possibly due to the nature of the instructions and task demands (Holbrook, Kawamoto, & Liu, 2017). Even if the segment is the minimal unit of articulation, participants might still strategically adopt differing response criteria to perform optimally given the assumed task demands. Of course, the fact that response criteria can vary down to a segment criterion – even if such a criterion is not always adopted by all participants for all tasks – is strong support for the segment as minimal unit of articulation.

With the segment as minimal unit of articulation and the adoption of a segment criterion, incremental articulation can occur within even a single monosyllabic word under the appropriate conditions (i.e., the onset is known while the coda is unknown). While this effect has been documented to occur in the naming task, the degree to which the segment acts as the minimal unit of articulation in other experimental tasks as well as in spontaneous speech is debatable. Critics of the segment as minimal unit of articulation have argued that the word naming task – which involves reading written words aloud – does not necessarily tap into universal speech production processes and therefore lacks external validity (e.g., Sulpizio & Burani, 2014). However, in Levelt’s model phonological encoding proceeds in the same way regardless of if the input phonemes are coming from the word-form
perception process or the non-orthographic lemma (Roelofs, 2004). According to this model, there is no reason to expect differing minimal units of articulation in word naming versus other naming tasks. Regardless, finding other empirical situations in which the segment acts as the minimal unit of articulation is useful for supporting the basic model framework as well as claiming the generality of the segment as minimal unit of articulation.

**Relation between Speech Production and Memory Recall**

Verbal memory recall can serve as a novel domain to investigate the effects of incremental articulation and the segment as the minimal unit of articulation. As spoken recall tasks involve producing speech without direct cuing or reading, they involve a greater degree of spontaneous production than naming tasks. Thus, evidence of the segment as the minimal unit of articulation in recall tasks would support and extend the previous findings. Before determining this though, it is important to examine the present research supporting an overlap between processes involved in verbal recall and speech production in order to justify applying findings from one field to the other.

**Phonological similarity effect.** The phonological similarity effect (PSE) arises when recalling a list of phonologically similar consonants or words in serial order. Items which are phonologically similar to each other can get sequential information confused, resulting in poor performance on those items. For example, the sequence of letters B, C, T, P, V is more confusable and results in lower serial recall accuracy than the sequence B, F, M, P, X (Conrad, 1964). The strict definition of
what counts as phonologically similar has varied somewhat through the literature, but
the evidence implies that such similarity occurs due to phonetic feature overlap
(Wickelgren, 1966; Mueller, Seymour, Kieras, & Meyer, 2003). More recently,
similarity effects have been shown to be due largely to acoustic similarity (defined as
manner of articulation), though articulatory similarity (defined as place of
articulation) contributes to the PSE when overt articulation of responses is required
(Schweppe, Grice, & Rummer, 2011).

The PSE has been thought to occur largely due to between-item interference
resulting in order confusion (Baddeley, 2007). Interestingly, when order is
unimportant to recall score, as in free recall, phonologically similar items tend to be
recalled at either the same rate or better than dissimilar items (Gathercole, Gardiner,
& Gregg, 1982; Poirier & Saint-Aubin, 1996). When phonologically similar items are
freely recalled, there are a greater number of shared cues allowing for redintegration
(reconstruction based on partial information) of multiple items; however, this
redintegrative process can also lead to multiple items being selected simultaneously
without regard for order, leading to the detrimental effects on serial recall.

Planning scope. Planning scope refers to how much of an utterance is
planned before the initiation of articulation. Planning scope can be variable, allowing
for more or less information planned before speech begins. The notion of response
criteria in the naming task is highly related to general planning scope, and as such is
related closely to the minimal unit of articulation – the smallest possible scope is one
minimal unit of articulation planned (as with a segment criterion), and scope can
increase in size and length from there. As planning scope is reduced, incremental articulation must increase to maintain fluid speech. For example, if only one word is fully planned upon initiation of articulation of a sentence, the remainder of the sentence must be incrementally articulated after the first word has been produced. If only one segment is planned upon initiation of articulation, the remainder of the word must be incrementally articulated.

Planning scope can also be conceptualized as prepared speech activated in working memory and awaiting production. In this view, units are held in working memory after they are encoded, allowing for the assembly and integration of higher level features, such as prosody within a full sentence and coarticulation between words. While the default planning scope is sometimes thought to be the phrase (Smith & Wheeldon, 1999; Martin, Crowther, Knight, & Tamborello II, 2010), there is also evidence for individual differences in default planning scope specifically related to working memory capacity as determined by verbal span (Swets, Jacovina, & Gerrig, 2014). Such findings explicitly link together working memory and planning scope.

**Chunking.** If information to be remembered forms larger chunks, such information is easier to remember. For instance, in chunking letters into words the unit of rehearsal (and therefore recall) has changed from the letter to a larger meaningful sequence (the word) stored in long-term memory (Cowan, 2005). Since chunking is the process of forming larger scale representations of small units, newly formed chunks themselves are now being treated as coherent and stable processing units. This has been found through measures of interresponse time (IRT) – the silent
gap in between recall responses – between chunked and unchunked items. IRT is
greater between unchunked items, suggesting that chunks are recalled – or
redintegrated – as complete packets (de Groot, 1965). Duration measures can then be
useful in providing information about chunking procedures, indicating what
information is chunked and the nature of the chunk size, as acoustic duration should
be stable after (but not before or during) full retrieval of a chunk.

Similar behavior can be seen in language processing, with individual
segments getting chunked into syllables that then get chunked into words. While
typically a full sentence is thought of a combination of various chunks – not a stored
chunk itself – larger units than the word such as multi-word phrases show frequency
effects (Arnon & Snider, 2010), indicating that phrases too are able to be chunked
from words and are represented in long-term memory. Such recursive representation
of information is quite interesting, and certainly a hallmark of expertise is the ability
to generate greater structure in smaller units by chunking them into larger pieces
made meaningful by stored representations in long-term memory (de Groot, 1965;
Cowan, 2005).

**Temporal Grouping.** While chunking relies on long-term semantic memory
representations, another form of grouping has been proposed that relies on temporal
relations. Stimuli grouped together in time (e.g., three sequences of three to four
numbers with pauses in between, as when reciting a telephone number) show
enhanced serial recall when compared to ungrouped stimuli under certain grouping
conditions (Hitch, Burgess, Towse, & Culpin, 1996). Since temporal groups
demonstrate a nested primacy effect – greater accuracy for the first item in a temporal group – it is thought that serial encoding is enhanced through a greater degree of disambiguation in position and item representations, and has been modeled as such (Burgess & Hitch, 1999).

A clear link is also available between temporal grouping and language processing. In language, the metrical structure of a word can be seen as a form of temporal grouping, as a multisyllabic word can be represented as a series of temporally grouped syllables (with stress included, which might also aid grouping; Boucher, 2006). More obviously, the prosody of a spoken sentence can also be seen as being temporally grouped, with intonation and rate of speech being cues to temporal grouping in working memory. This grouping then works to increase the effective capacity of working memory (Cowan, 2005), resulting in improved linguistic perception (e.g., binding clauses over long sentence distances). Thus, speakers can use prosodic cues to alleviate working memory load both on themselves (production) and on their listener (comprehension). While it is unclear if working memory capacity restricts language processing or if language processing works to develop working memory capacity in an optimal fashion, the connection between the two areas is well-established and intuitive.

**Chronometric Measures in Recall Tasks**

While accuracy is the preferred dependent variable in recall research, there have been efforts to incorporate chronometric variables into the understanding of memory dating back to work done by Sternberg et al. (1978), who had participants
recall sequences of common words (in overlearned sequences; e.g., days of the week, numbers) in a short-term memory recall task, varying the number of items in the sequence and the complexity (based on the number of syllables) of those items. They measured response latency – the time from the cue to begin recalling until acoustic onset – as well as response duration – the interval from acoustic onset until completed recall of the final word. Initial response latency was found to increase as the number of items as well as their complexity increased in the set. Crucially, they asked their participants to “complete their utterances rapidly” (Sternberg et al., 1978, p. 126), and this request was clearly met due to the incredibly fast rate of speech observed (over nine syllables per second on average). Thus, this corresponds to a minimal degree of incremental articulation, as the whole response should be planned out and ready before participants begin to respond; indeed, the presence of longer initial latencies for longer upcoming sequences supports the presence of a large planning scope whereby the entire sequence is planned in working memory before articulation begins. These results also imply that, given different task demands, the planning scope of verbal recall would be affected.

Measurements of latency are also useful in determining the time-course of a full response interval. By determining the latencies of each response in a sequence, the time between responses – IRT – can be quantified. Rohrer and Wixted (1994) used IRT in a verbal recall task to conceptualize the time-course of responding. However, there was one potentially large flaw with their study – the use of a voice key alone in determining latency. Voice keys are highly problematic as they are both
inconsistent between phonemic classes of initial segments (Kessler et al., 2002) and completely neglect the possibility of incremental articulation by assuming that acoustic latency signals the end of planning or encoding (Kawamoto, 1999), a highly problematic assumption given the complex relation and interactivity between articulation and ongoing cognitive processes (e.g., Fink et al., 2018). Specifically, response latency alone is unable to distinguish incremental articulation of a word from the case in which a word has been fully encoded but is not produced incrementally when both cases have the same latency. If incremental articulation is occurring within words, then IRT as determined by a voice key is not a valid measure of search termination and full item retrieval; at best, it can be quantified as a measure of articulatory initiation. However, the degree to which incremental articulation is occurring – if it is occurring at all – within a word during verbal recall tasks is at this point unknown and remains an empirical question.

In addition to latency measures, durations of words within a response have also been measured in the recall task. Word duration measures have shown that search processes likely overlap with articulation, an effect typically thought to manifest during a search for the subsequent word and not a search for the remainder of the currently articulated word (Haberlandt, Lawrence, Krohn, Bower, & Thomas, 2005). This means duration differences are not due to incremental articulation of the word currently being spoken, although this is more of an assumption than a fact. In the case of monosyllabic words, standard models of speech production (e.g., Levelt et al., 1999) imply that no articulation can occur before full encoding – and therefore
full retrieval – of the desired word. Of course, as already discussed the segment can indeed act as the minimal unit of articulation (Kawamoto et al., 2015), rendering this assumption incorrect. If this principle of speech production holds true during recall, it seems entirely possible that articulation of a phonologically overlapping segment or feature could occur while searching for the phonologically distinct remainder of the target word. For example, if all words in a list begin with /m/ such as mop, met, must, match, man, could /m/ be articulated before the remainder of the word is retrieved? If so, then segments could manifest the same duration effects during recall as seen in the naming task.

Can duration measures help to explain the phonological similarity effect? Although the effect is defined by reduced accuracy of phonologically similar items in a sequence (Conrad, 1964), the acoustic duration of the similar components of items might provide information about a possible cause of the PSE. Consider the definition of item interference from Cowan, Wood, Nugent, and Treisman (1997): “Stored information may be degraded by interference from ongoing processing or activity, at any stage of acquisition or retrieval […] As the duration of a spoken stimulus increases, the interference it produces increases. This increase is a function of the total auditory input to the phonological system, which is greater for words that take longer to utter” (p. 294). Thus, if linguistic units are elongated during articulation, the interference produced by them will be greater, resulting in poor recall, especially serial recall (as redintegrative processes might overcome interference during free recall). If this process manifests acoustically during verbal recall, then duration
measures could be used to quantify the degree of interference generated by phonemes in a word during production.

Chronometric methods for examining incremental articulation can also be used to investigate planning scope. The length of time to begin speaking as well as the duration of produced speech can be related to provide insight into planning scope. For example, Ferreira and Swets (2002) found that sentence production became more incremental (reduced planning scope) when participants were forced to begin their response before adequate information was given to plan the full response. Using duration measures, they found that participants elongated the end of the planned response while planning the remainder of their response under temporally demanding conditions (in this case, supplying the results of an addition task). When participants were able to comfortably plan the full response before speaking, word durations were shorter overall. Thus, planning scope can easily be measured through using response duration – all other factors being equal, longer duration of items within a full response equal a shorter initial planning scope.

Summary

The overlap between speech production and memory recall processes appear to be fairly clear, separated mostly by the dominant terminology and models. As such, results acquired from production studies should inform recall studies and vice versa. Most importantly, models of memory recall should not be considered as computationally isolated from linguistic processes as they have been in many models (e.g., Baddeley & Hitch, 1974; Levelt et al., 1999). Instead, the implications of the
PSE, planning scope, chunking and grouping show that memory recall processes are embedded within a larger framework of linguistic processes and long-term memory (Cowan, 1999; Hurlstone, Hitch, & Baddeley, 2014).

The existence of incremental articulation and the possibility of the segment as minimal unit of articulation can be exploited to further investigate the complex processes involved in memory recall tasks. In addition, providing evidence that the segment can act as the minimal unit of articulation in tasks other than the naming task can extend prior findings and reject the notion that this only occurs when reading single words aloud. If incremental articulation does indeed occur during recall tasks, measuring the form and degree of incrementality can serve as an additional measure of chunking behaviors, contribute to an explanation of the PSE, and explore the complex relation between retrieval, recall, and speech production.

**Experiment 1**

To determine the degree to which segment duration varies systematically in tasks other than the naming task, we investigated the difference in initial segment durations (ISD) in heterogeneous and homogeneous initial segment blocks using a standard recall paradigm, immediate free recall. Ideally, such a task encourages diverse strategies and grouping techniques, allowing for a broad investigation of any potential effect. For some blocks, all words to be recalled began with the same letter (homogeneous) and for the others the words began with a variety of letters (heterogeneous). Thus, in homogeneous blocks participants should always know the initial segment of the word they are trying to recall. What is unclear at this point is if
this knowledge will manifest behaviorally in the form of longer initial segment
durations in homogeneous blocks. To allow for this possibility, no plosive phonemes
were used, as plosives cannot be acoustically elongated and do not vary in acoustic
ISD (Kawamoto et al., 1998).

Also unknown is the degree to which elongated segments will be visible in
their actual acoustic production. Incremental articulation can also be quantified
through analyzing lip motion using video data (Kawamoto, Liu, Mura, & Sanchez,
2008; Kawamoto et al., 2014). Even in the absence of acoustic elongation,
participants might position their articulators to prepare for segment articulation, for
example opening the mouth and raising the tongue to prepare for an initial /n/,
rounding the lips to prepare for an initial /r/, or positioning the teeth over the lower lip
to prepare for an initial /f/.

It is important to consider that segment duration alone does not necessarily
indicate incremental articulation initiated solely by the initial segment. As has been
seen in both memory recall tasks (Sternberg et al., 1978; Haberlandt et al., 2005) and
investigations into incremental articulation (Ferreira & Swets, 2002), duration might
be affected not only by searching for the remainder of the current word, but searching
for the upcoming word as well. Specifically, in a task such as this, a long initial
segment duration might be part of a whole-word elongation produced while searching
for the subsequent word and not the remainder of the current word. This possibility
can be addressed by revising the dependent measure used. Instead of using a simple
measure of ISD, a proportional measure of ISD (ISD_{prop}) can be obtained by taking

16
the duration of the initial segment and dividing it by the duration of the entire word. This then provides a measure of percentage – what percent of a recalled word is composed of the initial segment alone. Although this measure might carry its own set of issues (which will be covered in the discussion), this is likely the most conservative way of estimating segment elongations while accounting for word elongations as well.

However, this only provides a measure of the word in recall, ignoring how different words are typically produced. In order to obtain the cleanest measure of segment elongations, ISD\textsubscript{prop} during recall can be compared to ISD\textsubscript{prop} during a standard reading task (the baseline measurement) to adjust for inter-item variability in ISD\textsubscript{prop}. This then provides a measure of ISD in proportion to word duration and adjusted for ISD during reading, and will be referred to as ISD\textsubscript{prop\_diff}.

**Method**

**Participants.** 14 undergraduates from University of California Santa Cruz participated for course credit. All participants were native English speakers and had normal or corrected to normal vision.

**Materials.** 100 monosyllabic words were chosen and used to generate ten lists of ten words each. All words had simple onsets and began with either \(F\), \(L\), \(N\), \(R\), or \(S\). These segments were chosen as they are all non-plosives and thus able to be acoustically elongated. They are also unambiguous in their pronunciation, mapping with one phoneme only. In addition, there was no overlap between these initial segments and the initial segments of fillers (e.g., \(M\) was not used as it would be hard
to distinguish from the fillers um and mm). All words had a vowel immediately after the initial segment, and the vowels were evenly distributed such that each possible orthographic vowel (five; A, E, I, O, U) appeared four times for each initial segment, making a total of 20 unique words beginning with each initial segment. While avoiding word-final plosive segments is ideal in order to obtain the most consistent measures of word duration between participants, the restrictions on words given above made it impossible to gather a sufficient number of stimuli for balancing purposes, and thus they were included. A complete list of the words chosen can be found in the Appendix.

The chosen words were then organized into lists of ten. Half of the words were sorted into homogeneous lists, with all ten words beginning with the same initial segment. The other half were mixed into heterogeneous lists, with each initial segment appearing twice. Every list of words contained no shared rimes, and phonological similarity within lists (aside from the initial segment) was attempted to be minimized. Beyond this, the word lists were assembled randomly. The condition a word appeared in (homogeneous or heterogeneous) was counterbalanced between participants.

A post-test was also constructed, consisting of two questions printed on a half sheet of computer paper. The first question was “Did you notice anything about the lists of words you studied?” and the second question was “Did you use any strategies to help you remember the word lists?”
**Apparatus.** The experiment was controlled by a Dell Optiplex 350 computer running ePrime experiment software, connected to a 22-inch Dell monitor with a refresh rate of 60 hz. The stimuli spanned approximately 1.1 to 2.5 degrees of visual angle from where the participants were seated.

Video data were recorded using a SONY EXview HAD 480TVL CCD lipstick camera. To minimize the effects of head movement on the video data, the lipstick camera was mounted 6 inches in front of a Wilson adjustable batting helmet, model A5240, on a 1/2 inch wide L shaped aluminum track bolted to the left side of the helmet. The video and audio capture were controlled by a Dell Optiplex GX260 desktop computer using Adobe Premiere Pro 1.5 in conjunction with an ADS PYRO A/V LinK IEEE1394a high-quality analog to digital video converter. All digitized video footage was captured at a frame rate of 30-frames/second with a 640 x 480 pixels resolution and an audio setting of a dual channel stereo at a 48-kHz sampling rate. The verbal responses were recorded using an Ikam lavalier microphone attached next to the camera and amplified to the proper level using a small Xenyx 302USB mixer.

**Procedure.** Participants were given a copy of the consent form, and provided informed consent about the nature of the experiment and data collection. The experimenter then painted four dots around the participant’s mouth (one above the lips, one below the lips, and one on each side of the lips) with Rubie’s brand blue cream makeup to assist with motion tracking. Participants were seated in front of the experiment running computer, and were asked to place a shower cap on their heads.
followed by the helmet camera. The experimenter then ensured that the lipstick camera was properly centered and able to see both the participant’s lips as well as their eyes. After making adjustments, the experimenter then read the instructions of the task aloud to the participant and asked if they had any questions. The participants then ran through two practice blocks before beginning the ten test blocks. The ten test blocks were presented in random order, with homogeneous and heterogeneous blocks intermixed randomly. The order of words within each list was also randomized.

Every block was structured in the same way. First, participants read the words in a given block one at a time as they were displayed on screen. Each word was displayed for one second and participants were instructed to read them with a “normal” pace and volume. After all ten words were presented once, a signal to respond was given (three asterisks) and participants recalled the list of words previously seen to the best of their ability. Participants were given ten seconds to recall as many words as possible. After ten seconds, the word “STOP” appeared on screen. Participants were instructed to finish any word they had already begun to speak, but to not begin a new word after the stop signal was presented. After the ten test blocks, participants were administered the post-test and debriefed as to the purpose of the experiment.

Results

Acoustic data were analyzed using PRAAT linguistic analysis software (Boersma, 2001). Duration measures for both words and segments were coded by hand in both the reading and recall portions of the task by trained research assistants.
Prior to analysis, responses were screened for errors. An error consisted of recalling any word not present in that particular block or mispronouncing a word (with respect to correct pronunciation during the reading portion). Data were also screened for verbal fillers (e.g., um) and non-linguistic responses (e.g., sighs), and these were not included in the following analysis. Boxplots for the raw measure of ISD during recall can be found in Figure 1.

ISD\textsubscript{prop} for both the reading and recall portions were determined by dividing the ISD of a correct word by the duration of the entire word. Baseline ISD\textsubscript{prop} measures were obtained by taking the means of ISD\textsubscript{prop} during reading after trimming extreme outliers (+/- 5% of the overall means). The final measure of difference, ISD\textsubscript{prop_diff}, was then calculated by taken the difference between ISD\textsubscript{prop} in the recall and reading portions. Thus, positive values of ISD\textsubscript{prop_diff} are obtained when proportional ISD was greater in the recall portion. In addition to providing an easy to understand measure, this also had the benefit of greatly reducing skew in the data, acting in a manner similar to a data transformation. Summary statistics for ISD and ISD\textsubscript{prop_diff} can be found in Table 1, and boxplots for ISD\textsubscript{prop_diff} across condition can be found in Figure 2.

As ISD\textsubscript{prop_diff} is still at least partially determined by the overall length of the word, it is likely that it varies greatly between words. In addition, exploration of the data revealed a potentially large difference in ISD\textsubscript{prop_diff} between participants. Since both are classified as random effects, the most appropriate statistical methodology was determined to be a crossed random-effects model (CREM). Given this, data were
analyzed using the general methodology outlined by Carson and Beeson (2013) for analyzing psycholinguistic data using a CREM. The model was constructed by adding factors in one at a time, and used a restricted maximum likelihood ratio of estimation.

A model which included the random effects of participants and items was tested first to determine the necessity of inclusion in the final model. The random effects of both participants ($Z = 2.184, p < .05$) and items ($Z = 2.346, p < .05$) were significant. The fixed effect of block condition was then added into the model, and was also significant ($F(1, 532) = 5.253, p < .05$), with homogeneous ISD$_{prop\_diff}$ ($N = 285$) significantly larger than heterogeneous ISD$_{prop\_diff}$ ($N = 271$). Thus, the final model included the random effects of participants and items as well as the fixed effect of condition. Summary parameters for the final model can be found in Table 2.
As ISD variability and differences between conditions were found with the acoustic data, the video data was deemed redundant for the purposes of the present experiment, and not coded and analyzed. Informal viewing of the video data appeared to show the predicted effect of articulator positioning to prepare for upcoming retrievals, but was determined to not provide enough additional information to warrant quantifying the data.

The post-tests were examined to provide information about what sort of strategies participants were using. One post-test was accidentally not administered, and could not be examined. The majority of participants (62%) indicated using some sort of semantic grouping strategy, such as linking words into “coupled pairs” or trying to “make a story”. Interestingly, the participant who produced the longest segment elongations did not indicate using this strategy. Participants also indicated

![Figure 2. Boxplot for ISD_{prop_diff} for both conditions in Experiment 1.](image)
using recency effects to their advantage (46%), which was clearly seen in the data as well (there was a high likelihood of the last word presented being recalled first). The strategy of one participant was particularly interesting: “For the lists where all the words started with the same letter, once I couldn’t remember automatically which words were shown to me, I would just try to think of all words I know that start with that letter to see if any were familiar.” While no other participants indicated using this particular strategy, it is an interesting method of attempting to turn the recall task into something more like a recognition task. (However, it did not seem to help too much, as they only recalled two additional words in homogeneous blocks over heterogeneous blocks.)
Discussion

Results from the experiment provided support for the existence of incremental articulation and the possibility of the segment as the minimal unit of articulation during recall tasks. ISD during recall was found to be significantly longer within homogeneous blocks than heterogeneous blocks, as was predicted. In addition, the distribution of recall ISD across blocks (Figures 1 and 2) showed more numerous and more extreme outliers in homogeneous blocks than in heterogeneous blocks. Although the effect size was fairly small when looking at mean differences, the differences in the distribution of raw scores provides a more direct picture of the variability between conditions and the nature of incremental articulation during this task.

ISD differences between the reading and recall portions were not due to elongation of the entire word. Elongation of the end of a word (opposed to the beginning) in this task would likely indicate a search for the next possible item to be retrieved (Sternberg et al., 1978; Haberlandt et al., 2005), and – while incremental articulation is still occurring in such a situation – this would not necessarily be due to adoption of the segment as the minimal unit of articulation. Through using ISD$_{\text{prop.diff}}$ instead of the raw measure of ISD, differences in whole word elongation were explicitly accounted for, and were found to not be responsible for the segment elongation effects seen through looking at ISD only.

Interestingly, there were trials in the heterogeneous condition in which participants appeared to be incrementally articulating their response. While this was
not strictly predicted, it was never conceptualized as impossible. The presence of initial segment elongation on heterogeneous trials simply indicates that, occasionally, participants were able to retrieve the first segment of one of the words in the block from a smaller set. In other words, in the heterogeneous condition two of the target words still began with the same segment as opposed to all target words in the homogeneous condition. All blocks also consisted of the same five initial segments, greatly restricting the possible initial segment options. As partial retrieval of words has been shown to exist during tip-of-the-tongue states (Brown & McNeill, 1966), retrieval and articulation of word-initial segments should also be possible even without phonologically overlapping initial segments. In fact, initial segment elongations in heterogeneous blocks (although less likely than in homogeneous blocks) are expected if partial retrieval is responsible for the effect. The greater frequency and duration of segment elongations in the homogeneous condition, as was originally predicted, is reflected in the mean difference as well as the greater proportion and durations of outliers in that condition.

One possible issue with the measurement used resulted from the use of words with word-final plosives. This is a problem when measuring the duration of the entire word, as word final plosives can either be produced as a separate phoneme or produced as a stop on the previous phoneme. This can result in a word duration difference between participants as well as between the reading and recall portion, and might help to explain the large effect of items seen in the mixed model analysis. While this was necessary to obtain enough stimuli, future studies should attempt to
avoid using words with word-final plosives when possible. Although possibly concerning, the similarity (aside from the transformation) between the boxplots of the raw ISD measure (Figure 1) and the adjusted ISD_{prop.diff} measure (Figure 2) shows that it did not have too large of an effect. If anything, the difference in pronunciation between portions would likely reduce ISD_{prop.diff} values, leading to an underestimation of effect size.

In summary, Experiment 1 answered some questions while raising many others. First, incremental articulation at the segmental level was found to occur in a task other than the naming task, showing the importance of measuring segment durations in a wide variety of experimental paradigms. The existence of incremental articulation in a more spontaneous task such as recall provides support for the segment as the minimal unit of articulation as a general phenomenon in speech production. Second, recall of words does not necessarily have to involve retrieval of a full word all at once. Some information about the currently searched for word (in this case, its initial segment) can be made available for articulation before the full word has been retrieved. While not strictly predicted, this concept is strengthened by the existence of segment elongations on heterogeneous blocks as well as homogeneous, as well as its greater degree and likelihood on homogeneous blocks. The presence of elongations in heterogeneous blocks needs to be explored further to reach a satisfactory explanation. This finding invites an alternative explanation for the cause of elongations during verbal recall.
As was originally conceived, elongations might occur as search occurs as the first segment is known and able to be articulated while the remainder of the word is being searched for. Alternatively, elongations might occur due to response competition and inhibition of previously studied material interfering with retrieval of the target item. While explaining elongations through search supports the use of the segment as minimal unit of articulation (as nothing is known about the upcoming response other than the initial segment, which is driving articulation), explaining elongations through interference does not necessarily support this position. If interference is the driving factor, a full candidate word might be known and encoded by the participant upon initiating articulation – the elongation is occurring due to delays in checking the candidate word against previously studied and reported words.

**Experiment 2**

The purpose of Experiment 2 was three-fold. First, as this is a novel finding and the sample size from Experiment 1 was relatively small, we wanted to ensure we could replicate the basic findings with a much larger sample. With this expanded dataset, additional analyses concerning block order, trial position, and errors can be examined. Most importantly, the order of blocks was controlled for, and not randomized. Half of the participants saw all the heterogeneous blocks first, and the other half saw all the homogenous blocks first.

With this design, the effects of interference can be more closely examined. Specifically, when seeing homogeneous blocks first (homogeneous portion) there should be no between-block interference based on the initial segment for those
blocks, and when seeing heterogeneous blocks first (heterogeneous portion) there should be less between-block interference when compared to seeing heterogeneous blocks last. This might be especially useful in helping to determine the cause of elongations seen previously in heterogeneous blocks – if between-block interference is contributing to elongations in heterogeneous blocks, ISD should be larger when the heterogeneous portion is presented last over first.

While looking at overall block number for the entire session can provide information about interference, with this design more detailed information can be obtained through looking at block number within a portion. When all homogeneous blocks are presented first, there should be no between-block interference based on the initial segment regardless of which number block it is in the portion. However, when all heterogeneous blocks are presented first, as the number of blocks in the heterogeneous portion increases, between-block interference based on the initial segment continuously increases. Given this, an interference account would also predict no effect of block number within the homogeneous portion and a positive correlation between ISD and block number within the heterogeneous portion.

Method

Participants. 60 undergraduates from University of California Santa Cruz participated for course credit. All participants were native English speakers and had normal or corrected to normal vision. Given the results from Experiment 1 and a desired power level of .80, at least 30 participants per condition were necessary.
**Materials and Apparatus.** The stimuli and materials, were the same as in Experiment 1. The recording apparatus was slightly altered due to failure of one of the technical components during the time in between the two experiments. The video and audio capture were controlled by a HP 500B desktop computer using ArcSoft video capture software in conjunction with a Top-Longer AV to USB converter. All digitized video footage was captured at a frame rate of 30-frames/second with a 720x480 pixels resolution and an audio setting of a dual channel stereo at a 48 kHz sampling rate. The remainder of the recording equipment was the same as in Experiment 1.

**Procedure.** The procedure was mostly the same as Experiment 1. In contrast to a fully random order, here either all heterogeneous blocks were shown first or all homogeneous blocks were shown first. Blocks within a condition were still randomized. The order of conditions as well as the words in each condition were counterbalanced between participants.

No post-test was administered after the experiment.

**Results**

Data were first analyzed in accordance with Experiment 1. As with Experiment 1, video data were not analyzed. All incorrect responses and verbal fillers (e.g., um, uh, or sighs) were removed from the data. The incorrect responses were put into a new dataset and analyzed separately. This resulted in 2427 correct responses across the 60 participants. Preliminary analyses from both Experiment 1 and the present Experiment suggested no difference in the pattern of results between
ISD\textsubscript{prop} and ISD\textsubscript{prop_diff}, and so ISD\textsubscript{prop} was chosen as the main dependent variable of interest due to simplicity of measurement and interpretation. The boxplot for ISD\textsubscript{prop} can be found in Figure 3, and summary statistics can be found in Table 3. Additional analyses for ISD\textsubscript{prop_diff} and reading and recall measures in isolation can be found in the Appendix.

**Proportional Measure.** A CREM was constructed using ISD\textsubscript{prop} as the dependent measure. Random effects were first added into the model. Both the random effect of participants (\(Z = 4.972, p < .001\)) and items (\(Z = 6.673, p < .001\)) were significant, and included in the following models. Fixed effects based on the predictions made (to minimize the number of parameters and models tested) were added into the model one at a time. The fixed effect of block condition was significant (\(F(1, 2320) = 20.822, p < .001\)), and included in the following models.
Table 3. Means and standard deviations for ISD$_{prop}$ for the relevant independent variables in Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Portion</th>
<th>Block Within Portion</th>
<th>ISD$_{prop}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>First</td>
<td>1</td>
<td>.2210 (.0837)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.2119 (.0828)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.2281 (.0802)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.2251 (.0892)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>.2241 (.0794)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2221 (.0831)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>1</td>
<td>.2189 (.0985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.2290 (.0825)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.2175 (.0886)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.2249 (.0989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>.2141 (.0842)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2207 (.0905)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2214 (.0866)</td>
</tr>
<tr>
<td>Homogeneous</td>
<td>First</td>
<td>1</td>
<td>.2359 (.1017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.2400 (.0967)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.2262 (.0962)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.2227 (.0906)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>.2313 (.1000)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2311 (.0969)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>1</td>
<td>.2272 (.0785)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>.2364 (.1057)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.2317 (.0926)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>.2307 (.1182)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>.2609 (.0785)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2369 (.0965)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>.2340 (.0967)</td>
</tr>
</tbody>
</table>

The effect of portion ordering was added next and found to be non-significant ($F(1, 2292) = .026, p = .872$), and not included. The interaction between block condition and portion ordering was also non-significant ($F(1, 62) = .075, p = .785$), and not included. The overall effect of block number within a portion was non-significant.
(F(4, 2294) = 1.525, p = .192), and not included. The overall interaction between block number within a portion and condition was non-significant (F(4, 2293) = .708, p = .586). The final model thus included the random effects of both participants and words as well as the fixed effect of condition, and can be seen in Table 4.

**Error Trials.** A great deal of information can be gathered by examining trials where participants made some sort of incorrect response. In total, there were 541 incorrect responses given, for a total error rate of about 18%, with more during heterogeneous blocks (N = 291) than homogeneous blocks (N = 250).

The most obvious instances of between-block interference resulting in error trials can be found through only looking at errors in which the word was presented on an earlier block. Such errors accounted for 53.5% of all error trials. Many other errors which were not strictly presented words were also highly related to presented words (e.g., nap instead of naps, or limp instead of lip). As such errors were the most common form of error, between-block interference is occurring quite frequently.

**Outcomes of Elongations.** The collected data can also be approached in a somewhat different fashion to provide additional information about what is occurring when participants elongate initial segments. The following analyses are thus focused only on trials with extreme segment elongations. Such trials were determined by taking only those trials with a difference in ISD greater than two standard deviations above the mean (ISD > 300ms) and RWD less than two standard deviations below the mean (RWD < 700ms). This resulted in a total of 50 correct extreme outlier trials being selected, with 16 from heterogeneous blocks and 34 from homogeneous blocks.
Table 4. Final models for $\text{ISD}_{\text{prop}}$ in Experiment 2. Bolded parameters are significant at $p < .05$. Parameter estimates are in seconds. $N_{\text{participants}}=60$; $N_{\text{items}}=50$. 

<table>
<thead>
<tr>
<th></th>
<th>ISD$_{\text{prop}}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>.004063 (.00012)</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>.001101 (.00022)</td>
<td></td>
</tr>
<tr>
<td>Items</td>
<td>.003283 (.00049)</td>
<td></td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.232863 (.00738)</td>
<td></td>
</tr>
<tr>
<td>Condition = HET</td>
<td>-.012672 (.00278)</td>
<td></td>
</tr>
<tr>
<td><strong>Model Summary</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance Statistic (-2LL)</td>
<td>-6066.457</td>
<td></td>
</tr>
<tr>
<td># of parameters</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

This was then compared with error trials, where participants made some sort of error (mispronunciation, repetition from same block, repetition from earlier block, or novel word recall). Extreme segment elongations were taken using the same method as above. This resulted in a total of 24 incorrect extreme outlier trials being selected, with 10 from heterogeneous blocks and 14 from homogeneous blocks.

Responses were also found where participants produced an initial segment only, without finishing production of an entire word. Within homogeneous blocks, all initial segment only productions were in line with the corresponding homogeneous segment. Filtering these trials under the same ISD restrictions as before resulted in 39 instances of such errors, with 8 occurring in heterogeneous blocks and 31 occurring in homogeneous blocks.

Combining this information, there were a total of 113 extreme segment elongations as defined here throughout the entire dataset, 34 in heterogeneous blocks and 79 in homogeneous blocks. In heterogeneous blocks, extreme segment elongations preceded correct responses 47.1% of the time, incorrect responses 29.4%
of the time, and no response (segment elongation only) 23.5% of the time. In contrast, in homogeneous blocks, extreme segment elongations preceded correct responses 43.0% of the time, incorrect responses 17.7% of the time, and no response 39.3% of the time.

**Recall Scores.** Although not the key dependent variable of interest, recall scores were also examined across conditions. Recall data were analyzed using a by-subjects repeated measures ANOVA, with block condition and order of blocks (homogeneous first or heterogeneous first) as fixed factors. Means and standard deviations can be found in Table 5. Recall scores were significantly higher in homogeneous blocks over heterogeneous blocks ($F(1, 1) = 41.326, p < .001$), and significantly higher when heterogeneous blocks were presented first over when homogeneous blocks were presented first ($F(1, 1) = 4.251, p = .044$). A significant interaction was also found between recall score and presentation order ($F(1, 1) = 4.480, p = .039$).

In addition to recall scores for correct trials, the error rate was analyzed in the same fashion. Error rates were calculated by taking each participants’ number of errors and dividing by the sum of the number of correct responses and number of errors. Means and standard deviations can be found in Table 5. Error rates were significantly higher in heterogeneous blocks over homogeneous blocks ($F(1, 1) = 8.796, p = .004$), and there was no difference in error rates when heterogeneous blocks were presented first and when homogeneous blocks were presented first ($F(1,$
No significant interaction was found between the two factors \( F(1, 1) = 1.371, p = .246 \).

Lastly, the correlations between recall scores and error rates with ISD\textsubscript{prop} were examined for both conditions to determine if segment elongations had any effect, positive or negative, on recall scores and error rates. No significant correlation was found between recall score and ISD\textsubscript{prop} for both heterogeneous blocks \((r(60) = .034, p = .795)\) and homogeneous blocks \((r(60) = .149, p = .256)\). In addition, no significant correlation was found between error rates and ISD\textsubscript{prop} for both heterogeneous blocks \((r(60) = -.217, p = .096)\) and homogeneous blocks \((r(60) = -.159, p = .224)\).

**Individual Differences.** The 60 participants who competed the task varied greatly in their production of pure segment elongations as defined above. Of the 60 participants, 18 produced at least one pure segment elongation and only 8 produced more than one. One participant in particular produced nine correct responses with extreme segment elongations during the experiment. Participants also varied greatly in their production of whole word elongations as determined by RWD > 700ms. Of the 60 participants, 25 produced at least one RWD elongation. Only two participants had more than three RWD elongations, and had 11 and 17 respectively. For these two participants elongating the entire word was much more common than elongating the initial segment only. Although both participants also showed a small number of extreme segment elongations as well, those occurred only in homogeneous blocks, while RWD elongations occurred in both conditions.
Table 5. Means and standard deviations of recall scores and error rates in Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Order</th>
<th>Recall Score</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>Heterogeneous First</td>
<td>20.10 (3.21)</td>
<td>.1898 (.101)</td>
</tr>
<tr>
<td></td>
<td>Homogeneous First</td>
<td>17.80 (3.22)</td>
<td>.2012 (.136)</td>
</tr>
<tr>
<td>Homogeneous</td>
<td>Heterogeneous First</td>
<td>21.93 (2.94)</td>
<td>.1660 (.093)</td>
</tr>
<tr>
<td></td>
<td>Homogeneous First</td>
<td>21.68 (2.97)</td>
<td>.1464 (.096)</td>
</tr>
</tbody>
</table>

Discussion

The results of Experiment 2 replicated and extended Experiment 1, explored the validity of multiple possible dependent variables as constructs of ISD, and provided some insight into the effect of interference on ISD.

The possibility that elongations and the overall ISD difference between conditions seen in Experiment 1 was due mainly to between-block interference was not directly supported. No significant effect of portion order or interaction between portion order and condition were found. Interestingly, in the key comparison heterogeneous words were trending in the opposite direction as was predicted (Figure 4). Regardless, neither the interaction nor the main effect of portion order was significant, and the potential effects of between-block interference were not found.

The effect of block number within a portion (homogeneous and heterogeneous portions) was also non-significant, with some evidence (see Appendix) of ISD being longest in the first block of a portion. As might be expected, the overall number of errors for the first block within a portion (94) was lower compared to the average number of errors on all other blocks (111.75), likely due to less between-block
Figure 4. Key comparison for determining overall effect of between-block interference in Experiment 2. Error bars represent one standard error of the mean.

interference within those first blocks which increases as the number of blocks seen within a portion increases. These results imply that between-block interference is having little effect on ISD, and what little effect it has is possibly resulting in shorter ISD as between-block interference increases from the first block to the remaining blocks.

Lastly, error trials in which initial segments were produced without accompanying word productions were found. Regardless of any additional analyses, these productions yield very compelling evidence for the use of the segment as minimal unit of articulation as they are likely indicators of a search which did not terminate in retrieval of a word. Considering participants’ overall willingness to produce incorrect responses during the recall portion, these productions are most
easily explained through search after retrieval of an initial segment. Comparing extreme segment elongations during correct retrievals with error trials contributed additional information into the nature of such elongations during recall. Extreme segment elongations were slightly more proportionally likely to result in an incorrect response or no response over an actual correct response.

Although significant in both Experiment 1 and 2, the between condition mean difference for ISD during recall was substantially larger in Experiment 1. This was likely due to both the smaller number of trials overall reducing the impact of extreme outliers as well as more extreme outliers in Experiment 1. In Experiment 1, the top three ISD outliers were all longer than 1.5 seconds (longer than any in Experiment 2), and all produced by the same participant. Because varying response criterion and planning scope among individuals can lead to these differences, it is not surprising to find that – due to random sampling – differences between participants within and across experiments. While many participants in both experiments produced at least one extreme segment elongation, none were as dramatic in raw length as this participant in Experiment 1.

The analyses of recall scores and error rates supported assumptions about the occurrence of interference. Specifically, the interaction between the order in which the blocks were presented and condition was in line with the initial justification for the experiment – recall scores were lower in the heterogeneous blocks when homogeneous blocks were presented first. However, homogeneous blocks were not affected in the same way – there was no difference in recall scores for homogeneous
blocks based on the order of block presentation. Alongside the fact that error rates were overall higher in heterogeneous blocks, it can be safely concluded that between-block interference is disproportionately affecting heterogeneous blocks over homogeneous blocks. This is clearly understandable – in homogeneous blocks the initial segments are not shared between blocks but in heterogeneous blocks they are, resulting in greater proactive interference in those blocks. The effect then does not show up in homogeneous blocks, as such between-block interference based on the initial segment is diminished as the number of homogeneous blocks seen increases and heterogeneous blocks are separated further back in time.

The present results provide some insight into why elongations occur in general, though many questions still remain. According to Sternberg et al. (1978), word elongations while searching for the next correct word in the set. While whole word elongations were clearly found, they were largely driven by ISD elongation and not RWD elongation – though RWD elongations did occur for many trials as well, they occurred alongside ISD elongations frequently. Importantly, ISD elongation and RWD elongation are not necessarily paired. While extreme RWD trials were highly positively correlated with ISD, extreme ISD outliers were not significantly correlated with RWD. This fact is also supported by the findings involving the main dependent measure, ISD$_{prop}$.

One possibility outlined earlier was that overtly articulating the initial segment of an unknown candidate word could result in greater redintegrative abilities and thus more success recalling correct words. As participants read the words aloud during
encoding, it might follow that overtly articulating segments during recall would have a noticeable impact on the success of recall due to cuing of context-relevant encoding factors (e.g., Tulving & Thomson, 1973). Given the present results, this appears to be a possibility. The 18 participants who produced at least one extreme segment elongation were found to have higher recall scores (M = 43.22, SD = 5.43) when compared to the 42 participants who did not produce any extreme segment elongations (M = 39.93, SD = 5.10), and this difference was significant while assuming unequal variances (t(30) = -2.192, p = .036). The total error rate was not significantly different while assuming unequal variance (t(45) = .909, p = .368). These findings of course conflict with the lack of correlations between ISD<sub>prop</sub> and recall scores for all participants across blocks as well as the finding that extreme elongations are slightly more likely to lead to errors, and thus should be interpreted carefully and likely result from additional individual differences, such as motivation.

**General Discussion**

The data from both experiments showed that segment elongations occur in the naming task independent of whole word elongations, segment elongations are more common and more extreme when the initial segment is homogeneous within a list, and segment elongations are not largely related to between-block proactive interference. The exact cause of and any potential strategic benefits to elongating segments remains unclear given the present results, and is worthy of future study. Regardless, the present results support and extend theories of speech production, support the segment as minimal unit of articulation, and highlight the benefit of
closely analyzing verbal response data in recall tasks with an eye to speech production findings in order to find novel empirical phenomena.

**Implications for Speech Production**

As shown in both experiments, incremental articulation can be easily quantified and explored in recall tasks. This is especially exciting as there is not yet much evidence for the segment acting as the minimal unit in tasks other than the naming task. As memory recall clearly does not involve reading segments, prior criticisms of the segment as the minimal unit of articulation as an artefact of the word naming task (e.g., Sulpizio & Burani, 2014) can be argued against based on the current evidence. The results thus provide support of the segment as minimal unit of articulation in general, and can be added to the growing body of literature supporting this position (outlined in Kawamoto et al., 2015).

Perhaps the most important reason to develop a strong theory of phonological encoding and speech production in general lies in the implications for research methodology and analysis in other areas of cognition. Response latency based on verbal productions is used as a common dependent variable across fields, and many methodological assumptions have been drawn from interpretations and theories about the nature of articulation and phonological encoding. If response latency is taken as a clear indicator of the termination of processing, the interactivity between articulation and cognition is being ignored (Fink et al., 2018). At best, ignoring this relation leads to missing possibly interesting effects; at worst, ignoring this information can lead to unjustified theoretical claims. Researchers utilizing verbal responses must at the very
least be mindful of the reality of the segment as minimal unit of articulation across varying speech production tasks when measuring acoustic latency.

**Implications for Memory Recall**

The results also convincingly show the irreconcilable problems with using voice keys or latency measures alone in general to measure IRT (Rohrer & Wixted, 1994). Similar to the issues with using a voice key with the naming task, it cannot be assumed that the first moment of acoustic production represents the termination of search. This is even more of an issue than originally thought, as it appears that incremental articulation based on the segment is possible even when all words in each block do not begin with identical initial segments. Through gathering measures of segment and word duration in addition to latency data, a more complete and accurate time-course of memory recall can be determined.

While the data delineate the phenomenon of segment elongation during recall quite nicely, it remains unclear as to what recall processes are driving this, as between-block interference seems to be having no effect. An additional possibility is that articulating the initial segment is being used intentionally to cue additional words for recall. As discussed above, such a strategy does not seem to be hugely effective, and is subject to a great degree of redintegrative interference. The unexpected finding that ISD was longest when in the first block of a condition (especially for the homogeneous condition) supports this possibility, albeit with some interpretation. It is possible that participants tried to use segment elongation as a cuing mechanism in early blocks – especially the first block – before realizing that it was not a great
strategy to retrieve additional words due to within-block interference; that is, articulating the initial segment was cuing words that had already been recalled in that block just as strongly as those that had not. Cuing was then somewhat abandoned as a useful strategy in later blocks, and re-evaluated when the condition type switched. This possibility also lines up with the lack of overall block ordering effects found and the lack of convincing between-block interference effects.

The results also provide some insight into potential mechanisms involved with the PSE, especially when considering the possible role of cuing as a driving force of segment elongation. As mentioned earlier, one possible explanation for the PSE is that interference is occurring during the rehearsal and recall stages between phonologically similar words (Cowan et al., 1997) leading to reduced performance on serial recall. Greater performance on phonologically similar lists in free recall – found in the present study as well – can be explained through redintegration and cuing based on phonologically similar portions. Given the current results however, this does not appear to be the case, as elongations were about equally likely to lead to a correct response or an incorrect response (including no complete response), and this did not vary greatly between heterogeneous blocks and homogeneous blocks.

In addition, the PSE benefit seen in free recall might also be due to contributions from recognition processes. As was mentioned by one participant in Experiment 1, recall of lists of phonologically similar items can be helped by considering all items that share the phonologically similar segment(s) and deciding if any are recognized. For example, in lists that all begin with /n/, search could occur by
retrieving any possible word that begins with /n/ and checking for recognition. Due to the additional requirement to maintain order in serial recall, turning the recall task into a recognition task would not have the same benefit, and would likely be detrimental – just because a word is recognized does not mean it is in the proper order.

On a more general note, the present results clearly show the variability of articulation during recall for both segments and entire words independently. According to the embodied perceptual-motor theory of short-term memory outlined by Macken, Taylor, and Jones (2014; also see Macken, Taylor, Kozlov, Hughes, & Jones, 2016), recall functions as retrieval of discrete gestures (be they written or articulatory) determined by object formation. As the present study specifically encouraged discrete object formation of each word in isolation (participants had to read the word aloud on the screen as it was presented in a “visual/vocal” condition; Macken et al., 2016), the degree of variability seen when producing the vocal response during recall does not easily fit in with this theory. Rather than supporting a conceptualization of recall in this manner, the results support more cognitive non-embodied models of recall with output processes resulting from search (e.g., Baddeley, 2007; Cowan, 1999).

Future Directions

While some claims about the PSE have already been put forth, additional research would help to confirm or deny the hypothesis put forth earlier. In order to best investigate the relation between ISD and the PSE, initial segment elongations can
be experimentally induced between blocks. This allows for a controlled examination of the relation between the two apart from individual differences. Participants could be presented with homogeneous blocks of words only and will be asked to articulate a single sound (either the shared initial segment or a phonological dissimilar segment; for the present set of stimuli, a vowel such as /ʌ/ would suffice) during the recall portion in between recall of words. If the PSE is due to interference during cuing – that is, serially inconsistent items all cued by the consistent homogeneous initial segment – then serial accuracy should be lower and free recall score should be higher when participants are forced to articulate corresponding initial segments during retrieval over inconsistent initial segments. Such results would support the possibility that between-item interference during rehearsal and retrieval results in (or at least contributes to) the PSE. Importantly, any observed differences would be occurring within the same exact lists of items, and thus not a quality of list similarity itself.

Future studies could also utilize ISD as a proxy for chunk or group recall. Since segment elongation manifests after production has begun but retrieval is still occurring, ISD differences should not appear when producing words within a retrieved temporal or semantic group or chunked sequence when retrieved as a group. Although not directly analyzed, participants in the present two experiments would often recall words in grouped sequences, for example recalling two or three words very rapidly at the beginning of recall – often the last words encoded, due to the recency effect, or the first words encoded due to the primacy effect – or recalling two words in very short durations when they were semantically linked (lining up with
participants self-reported semantic grouping strategy in Experiment 1). ISD could thus be used in a manner similar to how IRT has been used in the past, quantifying the start of chunk or group recall.

Video data, while collected, were not analyzed in either experiment. This was largely due to the success of acoustic ISD as a measure of incremental articulation making the lip positioning measure somewhat redundant. Some interesting effects – in addition to the predicted effects of lip positioning – were however determined through informal observation of the video data. Participants appeared to avert their gaze from the screen during recall, most often looking off to the upper right (from the participants’ point of view). Eye movements during recall have been studied with respect to visuospatial memory tasks (Tremblay, Saint-Aubin, & Jalbert, 2006; Martarelli & Mast, 2013). Specifically, retrieval of visuospatial information (e.g., grids of pictures of objects) is facilitated when participants fixate in the same area during recall as the objects were presented at during encoding (Johansson & Johansson, 2014). This benefit also appears to extend beyond visuospatial tasks to tasks involving recall of facts presented by virtual speakers in different areas of the screen (Richardson & Spivey, 2000). Participants were found to preferentially gaze in the area where the fact was presented to them when recalling the presented information. However, this process appears to be quite different as to what is occurring in the present recall task where participants intentionally avert their gaze from the area the word was presented in during encoding. As it would seem no systematic investigations into eye movements during simple verbal recall itself have
yet been conducted, this might be a fruitful avenue for further research to explore.

One possible theory is that looking away from the area where words were initially presented helps to lower interference, as all words were presented in the same location and would thus be cued to the same degree through maintaining gaze fixation in that location. If this is the case, then forcing participants to maintain gaze on the area of the screen where words were recalled should lead to reduced correct responses when compared to forcing participants to maintain gaze in an alternate area (perhaps not on the screen at all).

**Conclusion**

While not all speech production tasks explicitly involve recall and not all recall tasks explicitly involve speech, the amount of potential overlap between them is hard to ignore. Even tasks which appear to be cognitively isolated might rely on diverse cognitive abilities, and the study of such tasks might also be invigorated by incorporating methodology and theories from other areas. In this instance, measurements and analysis tools not typically thought of as a being a part of memory recall research have been shown to broaden our understanding of both respective processes. Through incorporating findings from speech production research into studies of verbal memory recall, both areas are benefitted and fruitful ideas for future studies can be generated.
Appendix A

Additional Analyses

Recall Measures

Additional models were built using the raw measures of ISD and remaining word duration (word duration minus ISD; RWD) during recall. As ISD$_{prop}$ is determined by both ISD and RWD, it remains important to ensure that the locus of differences in ISD$_{prop}$ is due to elongated ISD and not shortened RWD (since both would have similar effects on the calculated ISD$_{prop}$). For these analyses, the absolute measures of both ISD and RWD will be used, without comparing them to the reading portion or baseline reading durations.

First, an additional CREM for ISD was constructed. Random effects were first added into the model. Both the random effect of participants ($Z = 4.963, p < .001$) and items ($Z = 6.287, p < .001$) were significant, and included in the following models. Fixed effects were tested in the same way as above. The fixed effect of block condition was significant ($F(1, 2337) = 9.843, p = .002$), with homogeneous blocks having longer ISD than heterogeneous blocks (MD = .009, SE = .006) and included in the following models. Next, the effect of portion ordering was added and found to be non-significant ($F(1, 2286) = 1.577, p = .209$), and not included. The interaction between block condition and portion ordering was also non-significant ($F(2, 119) = .788, p = .457$), and not included. The overall effect of block number within a portion was significant ($F(4, 2293) = 3.839, p = .004$), and was included in the following models. The overall interaction between block number within a portion and condition was non-significant when both main effects were included ($F(8, 2292) = .346, p =$
.897). Similar to above, the effect of block number within a portion appears to be driven mainly by the difference between the first block presented compared to all other blocks, as determined by the significance of this parameter in the model ($t(2297) = 2.244, p = .025$) and its magnitude. However, unlike above, this effect was not due mainly to an interaction with condition. Interestingly, testing the simple effects revealed a result analogous to the above, with the overall effect of block number within a portion being significant in homogeneous blocks ($F(4, 2295) = 2.890, p = .021$) and not in heterogeneous blocks ($F(4, 2289) = 1.292, p = .271$). As the overall effect was non-significant, this interaction was not included in the final model. The final model thus included the random effects of both participants and words as well as the fixed effect of condition and block number within a portion.

Next, an additional CREM for RWD was constructed. Random effects were first added into the model. Both the random effect of participants ($Z = 5.158, p < .001$) and items ($Z = 6.225, p < .001$) were significant, and included in the following models. Fixed effects were tested in the same way as above. The fixed effect of block condition was non-significant ($F(1, 2338) = .836, p = .361$), and not included in the following models. Next, the effect of portion ordering was added and found to be significant ($F(1, 2285) = 9.399, p = .002$), and included in the following models. The interaction between block condition and portion ordering was non-significant ($F(2, 121) = .361, p < .698$), and not included. The overall effect of block number within a portion was significant ($F(4, 2294) = 3.878, p = .004$), and was included in the following models. Similar to above, the effect of block number within a portion
appears to be driven mainly by the difference between the first block presented compared to all other blocks, as determined by the significance of this parameter in the model ($t(2285) = 3.242, p = .001$) and its magnitude. The overall interaction between block number within a portion and ordering was non-significant when both main effects were included ($F(4, 2294) = 1.265, p = .282$). The final model thus included the random effects of both participants and words as well as the fixed effect of condition and block number within a portion.

**Reading Measures**

As data from all participants were also coded for their corresponding reading portions, it is possible to calculate $\text{ISD}_{\text{prop}}$ not from baseline reading proportions, but from everyone’s corresponding reading portion. This would then further remove variance from the model due to both idiosyncratic productions by participants as well as any consistent idiosyncrasies amongst the various research assistants (for this study, five in total) that coded initial segment onset, ISD, and word duration. However, this measure should be explored and validated before being used. If any systematic effects are occurring during the reading block, they could be mistaken as occurring during recall with the collapsed measure. In addition, effects occurring during the reading block are potentially interesting, as it is very similar to block primed naming tasks (e.g., Roelofs, 2002; Kawamoto et al., 2014).

Two models were built to examine ISD and remaining word duration (word duration minus ISD; RWD) in the reading portion alone. First, a CREM was constructed to examine ISD. Random effects of both participants ($Z = 5.170, p < .001$)
and items ($Z = 6.790, p < .001$) were significant, and included in the following models. Next, fixed effects were added in to the model separately, one at a time. The fixed effect of condition was significant ($F(1, 2298) = 4.208, p = .040$) with heterogeneous initial segments having slightly longer segment durations over homogeneous segments ($MD = .003, SE = .004$). The effect of portion order was non-significant ($F(1, 2276) = 1.260, p = .262$). The effect of block number within a portion was non-significant ($F(4, 2278) = 1.075, p = .367$). The interaction between condition and portion order was non-significant ($F(1, 63) = .005, p = .944$). The interaction between condition and block number within a portion was non-significant ($F(4, 2278) = 2.074, p = .082$), although it was trending in such a way that the simple effect in the heterogeneous condition was significant ($F(4, 2277) = 2.689, p = .030$). The final model thus included the random effects of participants and items and the fixed effect of condition.

Next, a CREM was constructed to examine RWD. Random effects of both participants ($Z = 5.199, p < .001$) and items ($Z = 6.690, p < .001$) were significant, and included in the following models. Next, fixed effects were added in to the model separately, one at a time. The fixed effect of condition was non-significant ($F(1, 2306) = .263, p = .608$). The effect of portion order was non-significant ($F(1, 2276) = .559, p = .455$). The effect of block number within a portion was non-significant ($F(4, 2281) = 1.494, p = .201$). The interaction between condition and portion order was non-significant ($F(1, 64) = 3.584, p = .063$). The interaction between condition and block number within a portion was significant ($F(4, 2280) = 3.712, p = .005$), with a
significant effect in the heterogeneous condition \( (F(4, 2279) = 2.948, p = .019) \) and a non-significant effect in the homogeneous condition \( (F(4, 2282) = 2.275, p = .059) \). Interestingly, this was not driven by a linear downward trend or by the first block in a portion as with the ISD\textsubscript{prop_diff} result. Parameter estimates showed the significant driving factor for this interaction was the difference in the fourth block in heterogeneous portion \( (t(2279) = -2.470, p = .014) \). The final model thus included the random effects of participants and items, the fixed effect of condition, and the interaction between condition and block number within a portion.

Using a participant specific method is concerning if assuming that participants’ reading productions are a stable baseline for which their recall should be compared. If the main goal of comparing with reading is to eliminate item variance in the recall condition, using a baseline measurement such as those calculated for ISD\textsubscript{prop_diff} is much more advisable. The systematic differences in variables of interest amongst the reading portion are interesting and could be explored in more detail. Particularly notable is the significantly shorter mean ISD for segments in heterogeneous blocks over homogeneous blocks, which appears to run counter to results from block primed reading studies (Kawamoto et al., 2014). Of course, the present methodology encouraged participants to “read the words in as natural of a manner as possible,” as opposed to the standard naming task instructions seen in block primed naming tasks, to “respond as quickly and as accurately as possible.”

**Proportional Baseline Measure**
A CREM was constructed using ISD$_{prop\_diff}$ as the dependent measure. Random effects were first added into the model. Both the random effect of participants ($Z = 4.993, p < .001$) and items ($Z = 3.120, p < .002$) were significant, and included in the following models. Fixed effects based on the predictions made (to minimize the number of parameters and models tested) were added into the model one at a time. The fixed effect of block condition was significant ($F(1, 2298) = 19.718, p < .001$), and included in the following models. Next, the effect of portion ordering was added and found to be non-significant ($F(1, 2331) = .043, p = .837$), and not included. The interaction between block condition and portion ordering was also non-significant ($F(2, 119) = .069, p = .934$), and not included. The overall effect of block number within a portion was non-significant ($F(4, 2355) = 1.941, p = .101$), and not included. The overall interaction between block number within a portion and condition was non-significant ($F(8, 2353) = 1.387, p = .197$); however, univariate tests revealed a significant effect of block number within the homogeneous condition ($F(4, 2356) = 2.420, p = .046$), and not the heterogeneous condition ($F(4, 2350) = .353, p = .842$). This effect appears to be driven mainly by the difference between the first block presented in the homogeneous condition when compared to all other blocks in the homogeneous condition, as determined by the marginal significance of this parameter in the model ($t(2363) = 1.814, p = .070$) and its magnitude. As the overall effect was deemed non-significant, this was still omitted from the final model. The final model thus included the random effects of both participants and words as well as the fixed effect of condition.
Given the previously determined issues with using participants own reading times as a baseline measurement, the ability to factor in item effects in mixed models, and given that results for ISD$_{prop,\text{diff}}$ were essentially the same as for ISD$_{prop}$, the simpler measure appears to be a perfectly valid dependent variable to use in future experiments.
Appendix B

Additional Summary Statistics and Models for Experiment 2

Table 6. Means and standard deviations in milliseconds for recall times (ISD and RWD) and as a proportional difference for ISD\textsubscript{prop\_diff} for the relevant independent variables in the additional analyses for Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Portion</th>
<th>Block Within Portion</th>
<th>ISD_Recall</th>
<th>RWD_Recall</th>
<th>ISDprop_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>First</td>
<td>1</td>
<td>142 (65)</td>
<td>493 (110)</td>
<td>0.0111 (.0611)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>128 (64)</td>
<td>464 (117)</td>
<td>-0.0008 (.0579)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>135 (67)</td>
<td>449 (109)</td>
<td>0.0132 (.0588)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>131 (65)</td>
<td>443 (113)</td>
<td>0.0174 (.0696)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>136 (57)</td>
<td>470 (107)</td>
<td>0.0112 (.0604)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>134 (64)</td>
<td>464 (112)</td>
<td>0.0105 (.0618)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>1</td>
<td>137 (107)</td>
<td>472 (161)</td>
<td>0.0108 (.0762)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>134 (62)</td>
<td>444 (94)</td>
<td>0.0118 (.0603)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>129 (84)</td>
<td>443 (113)</td>
<td>0.0060 (.0736)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>128 (68)</td>
<td>439 (127)</td>
<td>0.0069 (.0700)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>126 (63)</td>
<td>448 (106)</td>
<td>0.0045 (.0638)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>130 (78)</td>
<td>449 (122)</td>
<td>0.0079 (.0687)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>133 (71)</td>
<td>457 (117)</td>
<td>0.0093 (.0651)</td>
</tr>
<tr>
<td>Homogeneous</td>
<td>First</td>
<td>1</td>
<td>149 (102)</td>
<td>464 (123)</td>
<td>0.0367 (.0914)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>157 (132)</td>
<td>474 (158)</td>
<td>0.0109 (.0764)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>136 (75)</td>
<td>461 (123)</td>
<td>0.0198 (.0806)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>137 (80)</td>
<td>471 (178)</td>
<td>0.0109 (.0766)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>133 (69)</td>
<td>436 (112)</td>
<td>0.0141 (.0758)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>142 (94)</td>
<td>462 (142)</td>
<td>0.0185 (.0807)</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>1</td>
<td>139 (64)</td>
<td>465 (124)</td>
<td>0.0298 (.0689)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>144 (101)</td>
<td>444 (117)</td>
<td>0.0245 (.0904)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>140 (73)</td>
<td>457 (140)</td>
<td>0.0138 (.0651)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>130 (73)</td>
<td>441 (105)</td>
<td>0.0231 (.0862)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>161 (76)</td>
<td>445 (105)</td>
<td>0.0258 (.0707)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>142 (79)</td>
<td>451 (123)</td>
<td>0.0234 (.0770)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>142 (87)</td>
<td>456 (133)</td>
<td>0.0209 (.0788)</td>
</tr>
</tbody>
</table>
Table 7. Final models for ISD and RWD during recall, and ISD$_{prop\_diff}$ in additional analyses for Experiment 2. Bolded parameters are significant at $p < .05$. Parameter estimates are in seconds. N$_{participants}$=60; N$_{items}$=50.

<table>
<thead>
<tr>
<th></th>
<th>ISD Recall</th>
<th>RWD Recall</th>
<th>ISD$_{prop_diff}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>.003857 (.00012)</td>
<td>.008763 (.00026)</td>
<td>.004062 (.00012)</td>
</tr>
<tr>
<td>Participants</td>
<td>.001006 (.00020)</td>
<td>.003720 (.00072)</td>
<td>.001103 (.00022)</td>
</tr>
<tr>
<td>Items</td>
<td>.001496 (.00024)</td>
<td>.003129 (.00050)</td>
<td>.000135 (.00004)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.139420 (.00647)</td>
<td>.447648 (.01082)</td>
<td>.020415 (.00479)</td>
</tr>
<tr>
<td>Condition = HET</td>
<td>-.008565 (.00270)</td>
<td>-.011892 (.00268)</td>
<td></td>
</tr>
<tr>
<td>Portion = FIRST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BlockWithinPortion = 1</td>
<td>.012155 (.00411)</td>
<td>.020090 (.00620)</td>
<td></td>
</tr>
<tr>
<td>BlockWithinPortion = 2</td>
<td>.004685 (.00412)</td>
<td>.011132 (.00620)</td>
<td></td>
</tr>
<tr>
<td>BlockWithinPortion = 3</td>
<td>-.000716 (.00411)</td>
<td>.003958 (.00619)</td>
<td></td>
</tr>
<tr>
<td>BlockWithinPortion = 4</td>
<td>-.001107 (.00411)</td>
<td>.000267 (.00620)</td>
<td></td>
</tr>
<tr>
<td><strong>Model Summary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance Statistic (-2LL)</td>
<td>-6230.192</td>
<td>-4219.009</td>
<td>-6270.586</td>
</tr>
<tr>
<td># of parameters</td>
<td>9</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>
### Appendix C

**Stimuli Used**

<table>
<thead>
<tr>
<th>FACE</th>
<th>LAB</th>
<th>NAB</th>
<th>RACK</th>
<th>SALT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FADE</td>
<td>LAMP</td>
<td>NAG</td>
<td>RAIN</td>
<td>SAME</td>
</tr>
<tr>
<td>FANG</td>
<td>LAND</td>
<td>NAME</td>
<td>RARE</td>
<td>SASH</td>
</tr>
<tr>
<td>FARM</td>
<td>LASH</td>
<td>NAPS</td>
<td>RAT</td>
<td>SAVE</td>
</tr>
<tr>
<td>FEAR</td>
<td>LEARN</td>
<td>NEAR</td>
<td>REACH</td>
<td>SEEM</td>
</tr>
<tr>
<td>FEE</td>
<td>LEASH</td>
<td>NET</td>
<td>REEF</td>
<td>SELL</td>
</tr>
<tr>
<td>FENCE</td>
<td>LEFT</td>
<td>NEWS</td>
<td>RENT</td>
<td>SERVE</td>
</tr>
<tr>
<td>FERN</td>
<td>LENS</td>
<td>NEXT</td>
<td>REST</td>
<td>SET</td>
</tr>
<tr>
<td>FILM</td>
<td>LIFE</td>
<td>NICE</td>
<td>RIDGE</td>
<td>SICK</td>
</tr>
<tr>
<td>FIN</td>
<td>LIME</td>
<td>NIGHT</td>
<td>RIGHT</td>
<td>SIDE</td>
</tr>
<tr>
<td>FINE</td>
<td>LINK</td>
<td>NIGHT</td>
<td>RING</td>
<td>SIGN</td>
</tr>
<tr>
<td>FISH</td>
<td>LIP</td>
<td>NINE</td>
<td>RIP</td>
<td>SIT</td>
</tr>
<tr>
<td>FOLD</td>
<td>LOCK</td>
<td>NOISE</td>
<td>ROAR</td>
<td>SOAR</td>
</tr>
<tr>
<td>FOOD</td>
<td>LONE</td>
<td>NOON</td>
<td>ROAST</td>
<td>SOOT</td>
</tr>
<tr>
<td>FORK</td>
<td>LOST</td>
<td>NOSE</td>
<td>ROLL</td>
<td>SOUL</td>
</tr>
<tr>
<td>FORT</td>
<td>LOVE</td>
<td>NOW</td>
<td>ROT</td>
<td>SOUTH</td>
</tr>
<tr>
<td>FULL</td>
<td>LUCK</td>
<td>NULL</td>
<td>RUG</td>
<td>SUIT</td>
</tr>
<tr>
<td>FUND</td>
<td>LUMP</td>
<td>NUMB</td>
<td>RUNE</td>
<td>SUM</td>
</tr>
<tr>
<td>FUSE</td>
<td>LUNCH</td>
<td>NURSE</td>
<td>RUNG</td>
<td>SUN</td>
</tr>
<tr>
<td>FUSS</td>
<td>LUNG</td>
<td>NUT</td>
<td>RUT</td>
<td>SURF</td>
</tr>
</tbody>
</table>
References


Kawamoto, A. H. (1999). Incremental encoding and incremental articulation in speech production: Evidence based on response latency and initial segment


