Factors in Word Duration and Patterns of Segment Duration in Word-initial and Word-final Consonant Clusters

Becca Schwarzlose

1. Introduction

The words *playing* and *splaying* have many apparent similarities. In fact, they appear to be identical except that one of the words has an additional segment (/s/) in the first syllable. Therefore, one might expect the duration of the word *splaying* in a normal utterance to differ from the duration of the word *playing* only by the duration of /s/. We will call this line of reasoning the Additive Theory of Word Duration because it predicts that as one adds additional segments to a word (such as the progression from *laying* to *playing* to *splaying*), the durations of the individual segments remain stable and the durations of the words increase by the amount of time required for the new segment. We can think of this theory in terms of a stack of blocks. If you have a stack of blocks with a total height h and you add a block of height x to the stack, then the resulting pile will have a total height $h + x$. The addition of the new block has not changed the heights of the blocks that were already in the stack.

Another possible view of word duration offers very different predictions for such circumstances. According to this theory, speakers want to produce words that are isochronous. That is, they want their words to have equal durations. We will call this theory the Isochronal Theory of Word Duration. An important aspect of this theory is the necessary reduction of segment duration. If the durations of *playing* and *splaying* are the same when uttered, then the durations of some or all of the segments in *splaying* must be shorter than their counterparts in *playing* in order to compensate for the time required to pronounce the additional segment. Our stack of blocks cannot be used to illustrate this theory. Rather, we can use the metaphor of a crowded closet. There are so many clothes on hangers in this closet that they fill the closet from wall to wall. When you want to hang up another piece of clothing, you have to push the other items of clothing to one side. This compresses the items of clothing so that they take up less space and as a result you have room to hang your new sweater on the rod. In other words, all of your clothes (the segments) would fit beautifully into the closet (word) if it were just a little wider (had a greater duration), but since it is not, one must push the clothes together (shorten the durations of the segments) so that they take up less space and there is room to add one more.

The metaphors of the blocks and the closet are, of course, not meant to mirror the actual processes by which we form words; the lexical entry for *playing* would have a unified representation of the word rather than the instructions: take the word *laying* and add a *p*. However, in this case we will be tracking the durations of segments within words that systematically differ only in the presence or absence of one or two additional segments. For the purposes of studying how the durations of these segments vary in the contexts of these different words, it is useful for us to think in terms of adding segments, so long as we don t forget that this is a mental heuristic rather than a psychological reality. This said, we can move on to a description of the study that was performed in order to test these two theories of word duration.

2. Method

Materials

Three sets of words (totaling 17) were compiled in order to test the Additive and Isocronal Theories of Word Duration The *splaying* words had varied word-initial syllable onsets: *saying, paying, laying, playing, spaying, slaying, and splaying*. The *basks* words and *clasps* words varied in their codas: *bass, back, bask, backs, basks, class, clap, clasp, claps, and clasps*. Therefore, the *splaying* words had various consonant clusters in the first syllable onset while the *basks* and *clasps* words had various consonant clusters in their codas. This study focused on total word duration and the durations of segments in these consonant clusters. To limit the scope of the topic covered, vowel durations were not individually measured or analyzed. Moreover, the durations of the approximant /l/ in *splaying* consonant clusters was not analyzed because of the difficulty of accurately distinguishing it from the vowel that follows. Many of the target words had very different meanings and sometimes came from different lexical categories. As a result the sentences that provided context for these words were created to be vague or ambiguous in order to downplay lexical differences. (See Appendix 1 for the list of sentences.) Moreover, sentences were created with the goal of making word boundary identification easier and more accurate. Therefore the segments that directly preceded or followed the target word were easily distinguished from those belonging to the beginning or end of the target word. Another requirement of the sentences was that they neither start nor end with the target word. All words from a particular word group were presented in identical sentence contexts, except for *clasp* words. Since some of these nouns were singular and others were plural, two versions of the sentence were created. However, both versions contained the same number of syllables and in both versions the target word was preceded by a schwa. The sentences with a particular target word group were listed together, but in no meaningful order within the group. They were listed together so that subjects would be more likely to be speaking at a comparable rate for all of the sentences in that group. In order to resolve any confounds that could arise from sentences at the beginning or end of the list, the first sentence and the last two sentences in each group did not contain a target word, but rather a word somewhat similar to the target words.

Procedure and Subjects

Two men, ages 64 and 27, and one woman, age 20, participated in this experiment. The three subjects were members of the same immediate family: a father, a son, and a daughter. Each was given the list of sentences and was instructed to read the sentences aloud in a normal tone of voice, taking care to repeat any sentences in which they misspoke.

Using PRAAT software, the target word and segment boundaries were then identified. This was done both visually (by looking at the speech wave forms of the subjects utterances) and aurally. Cues such as periodicity and amplitude, as well as aural evaluations of timbre, were used to determine the locations of these boundaries.

3. Results and Discussion

Despite the fact that the subjects were members of the same family, they showed substantial variation in their rates of speaking. The father had an average target word duration of 664 ms, while the son averaged 586 ms per word and the daughter averaged 467 ms per word. Despite this variation in speaking rate, the percentages of segment durations out of the total durations of target words were similar across subjects. Moreover, the patterns found when word and segment durations were averaged across the three speakers (discussed below) were also apparent for the word and segment durations of each of the three speakers individually.

The mean expected word duration for each target word across subjects, as predicted by the Additive Theory, was found by calculating the expected word duration for each individual subject and averaging across the three subjects. The expected word duration for each individual subject was calculated by adding the duration of the additional segment when it appears alone in the onset or coda to the duration of the preexisting word. For example, the expected word duration of *playing* for Subject 1 would be calculated by adding the duration of /p/ in the word *paying* to the duration of the word *laying*. This was done for all three subjects and the mean values across subjects were then calculated for each word. Words within each word group were then grouped by the number of segments in their word-initial onsets (for *splaying* words) or by the number of segments in their codas (for *basks* and *clasps* words) and the mean duration was calculated for each group. The results are shown in Figures 1, 2, and 3. Since the expected value was calculated using the durations of the additional segments when they appeared alone in the onset or coda, the expected and observed values for words with a single segment in the onset or coda were equal by definition. According to the Additive Theory, we can predict that the expected values for words with two or three segments in their onsets or codas will be approximately equal to the observed values for each subgroup and that these values would increase as the number of segments in the onset or coda increases. The Isochronal Theory of Word Duration would predict that the observed values will be equal across number of segments in the onset or coda and that these values will be much less than the expected values. Figures 1, 2, and 3 do not completely uphold either of these predictions. As the Additive Theory predicts, the observed word durations increased as the number of segments in the onset or coda increases. However, the observed durations for words with 2 or 3 segments in the onset or coda were uniformly less than their expected durations, as was predicted by the Isochronal Theory. Figures 1, 2, and 3 illustrate what happened as the number of segments in the onset or coda increased: the observed durations increased, but they did not increase in a fully additive fashion. This suggests that speakers strive for isochrony but are unable to produce fully isochronal words, perhaps because of listener-oriented constraints. We will call this explanation the Flawed Isochronal Theory. A second possible theory differs from the Flawed Isochronal Theory in terms of motivation. According to this theory, which we will call the Minimal Effort Theory of Word Duration, speakers create shorter words than the Additive Theory predicts because speakers are motivated by a desire to minimize the effort and time required to produce words rather than by a desire to produce words of the same duration.

If word duration is not determined by the simple addition of segments of set duration, then the duration of segments must vary as the number of segments changes. For example, *splaying* did not have the duration predicted by the Additive Theory, we must assume that of some or all of the segment durations in *playing* are reduced (as is the duration of the new segment /s/) when /s/ is added to *playing*. Now that we have determined that the durations of some or all segments are reduced with the addition of a new segment, we must investigate the pattern of reduction among segments. A logical pattern would be that all segments lose time uniformly, so that the same amount of time is subtracted from the duration of every segment in the word in order to compensate for the added duration of the new segment. In Figure 4, we see the mean durations of the segment /s/ in *splaying* words beginning with /s/, grouped by the number of segments in the onset or coda. If the durations of all segments are uniformly reduced as the number of segments grow, then we would expect the duration of /s/ to decrease as the number of segments in the onset increases. Figure 4 shows that there is no such decrease for the duration of /s/ in *splaying* words. This fact argues against the theory that all segment lose time uniformly. Rather, it seems that some segment durations may be reduced more than others and that the durations of some segments are not reduced at all.

If segments do not lose time uniformly with the addition of new segments, then what factors determine which segments will lose time and to what degree? Perhaps the natural class of a segment determines the extent to which its duration is shortened. It is possible that the /s/ of *splaying* in Figure 4 is not shortened because it is an obstruent or a fricative or because it is voiced. Another possibility is that the position of the segment in the word determines whether the segment will be shortened. These two hypotheses can be tested by looking at two target words pairs /bQks/ - /b Qsk/ and /klQps/ - /kl Qsp/. The words in each pair contain exactly the same number of segments and the same individual segments. They differ only in the order of the consonants in the coda. (The *splaying* words did not include pairs of this kind and therefore could not be included in this particular analysis.) If the natural class of a segment determines the extent to which its duration is reduced, then we would expect the duration of the given coda consonant to remain constant for both words in a given pair. Figures 5 and 6 illustrate the average durations of the coda consonants for the four words. Clearly, these two graphs illustrate that the extent to which a segment s duration is reduced is not determined by its natural class, but rather by its position in the consonant cluster.

If position is the most important factor in determining compensatory segmental duration reduction, then we would expect segments in particular positions to consistently show less duration reduction than those in other positions. We can evaluate this hypothesis by examining Figures 7 and 8. The results from *splaying* words with two or more segments in their onsets were included in Figure 7 and Figure 8 contains data from *basks* and *clasps* words with two or more segments in their codas. Segments that occurred in the first or second position in the onset were studied for Figure 7 and segments in the last or second to last position in the coda were studied for Figure 8. For each segment measured in these contexts, its duration in the cluster context was divided by its duration when it appeared alone in the onset (for Figure 7) or in the coda (for Figure 8). In this way, the natural or inherent durational differences between different types of segments were taken into account and the resulting values reflected the duration of a segment relative to its duration when it is presumably not reduced. As a result, these values could be averaged across segment types without introducing potentially confounding variables. The result is two graphs that strongly support a model in which position in a consonant cluster determines extent of durational reduction. Moreover, Figure 7 shows that segments in the second position undergo much greater durational reduction than do those in the first position of the word-initial onset and Figure 8 illustrates the extent to which the durations of segments in the second to last position in a word-final coda are more greatly reduced than those of segments in the last position.

Overall, the results of this experiment indicate that words containing consonant clusters have shorter durations than we would expect if segments were pieced together in an additive fashion. At the same time, however, we have seen that such words are not isochronous; their durations do increase as the number of component segments increases. Since word duration is less than the value predicted by the Additive Theory, segments within the word must have shorter durations in order to compensate (in the same way that the clothes in the closet must take up less space in order for you to add a new sweater.) Furthermore, the results of this experiment suggest that the extent of the reduction of a segment s duration due to the addition of a new segment is not uniform across segments in the word, nor is it primarily determined by the natural class of the segment. Rather, the position of the word is what determines the extent of duration reduction. The duration patterns of segments in four possible positions in a word have been studied and we have found that the first and last segments in a word generally undergo little to no duration reduction as the number of segments in the word increases, while the second-to-first and the second-to-last segments show a substantial reduction in duration in such circumstances. This phenomenon could possibly be explained in terms of listener-oriented constraints. It may be important that the first and last segments maintain their original, longer durations because they may serve as markers that help to indicate the beginning and the end of a word for a listener. We will call this proposed constraint the Well-Marked Boundary Constraint.

There are many ways in which these investigations into patterns of word and segment duration reduction could be expanded upon. This study did not examine the patterns of vowel durations or the patterns of any other segments that are not in word-initial or word-final consonant clusters; such inquiries would be essential to a complete picture of durational change patterns as the number of segments increase. Furthermore, to be secure in the statement that natural class plays no role in deciding the level of duration reduction, further studies should investigate whether the effects found here for voiceless fricatives and plosives apply equally to voiced obstruents, approximants, and nasals. Finally, whole words with word-initial, word-final, and word-medial consonant clusters should be measured for their overall durations and the durations of every one of their component segments. This will allow a more complete investigation of the global patterns of segment duration reduction in words as new segments are added.

In the absence of additional data from the suggested studies above, we are left with two distinct theories that could equally account for our results. One of these theories is the Flawed Isochronal Theory of Word Duration described above. According to this theory, speakers strive to produce words with equal durations. However, they are unable to fully succeed at this because of listener-oriented constraints. If segmental durations were reduced beyond a certain extent, listeners would be likely to have difficulty correctly identifying all the segments in a word. Furthermore, the proposed Well-Marked Boundary Constraint may require that the first and last segments undergo minimal or no duration reduction, thereby adding another obstacle in this quest for isochrony. The Minimal Effort Theory (also described above) suggests instead that speakers reduce the durations of all segments as much as possible without making their speech unintelligible to listeners in order to minimize the effort and time that is dedicate to speaking.

Listener-oriented constraints, such as the proposed Well-Marked Boundary Constraint, would be influential in determining the patterns of segment duration under this theory as well. Both the Flawed Isochronal Theory and the Minimal Effort Theory could account for the results of this study, as they would both result in minimal durations for those segments in positions for which listener-oriented constraints do not intervene. Further thought and investigation will be required in order to identify which theory is in play in the reduction of segment durations or if perhaps both the maximization of isochrony and the minimization of effort are motivating forces behind this phenomenon.

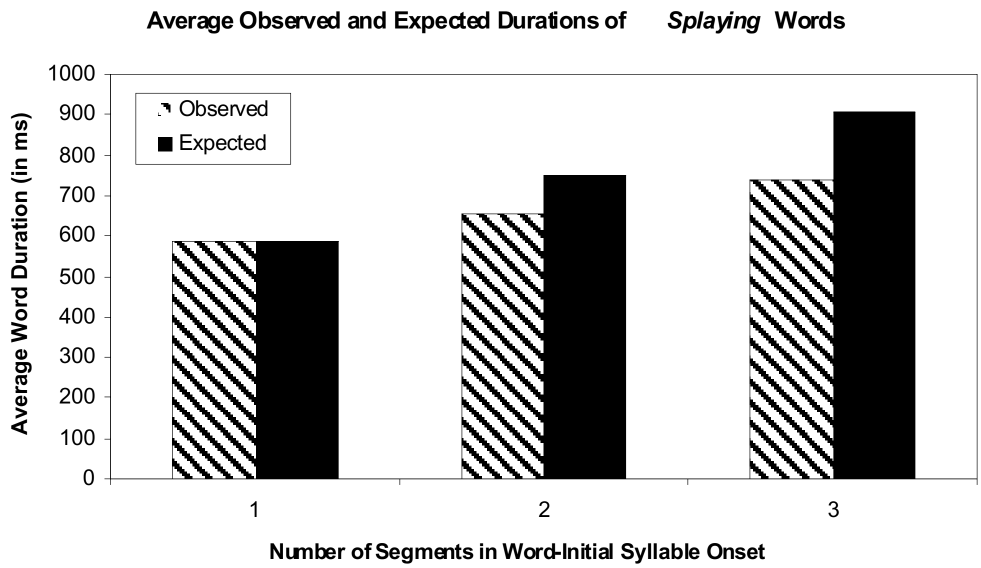

Figure 1. The above figure depicts the differences between the mean observed durations for *splaying* words and the mean durations we would expect based on our additive theory. The method used to calculate the mean expected values is described in the Results and Discussion section. The x-axis represents the number of segments in the word-initial syllable onset and the y-axis denotes average word duration in milliseconds. The observed and expected values for words with one segment in the onset are equal by definition. Since the observed values are less than the expected values when there are two or three segments in the onset, this figure provides evidence contrary to the Additive Theory. However, the observed values are not equal across onset consonant number, indicating that the words are not perfectly isocronal either.

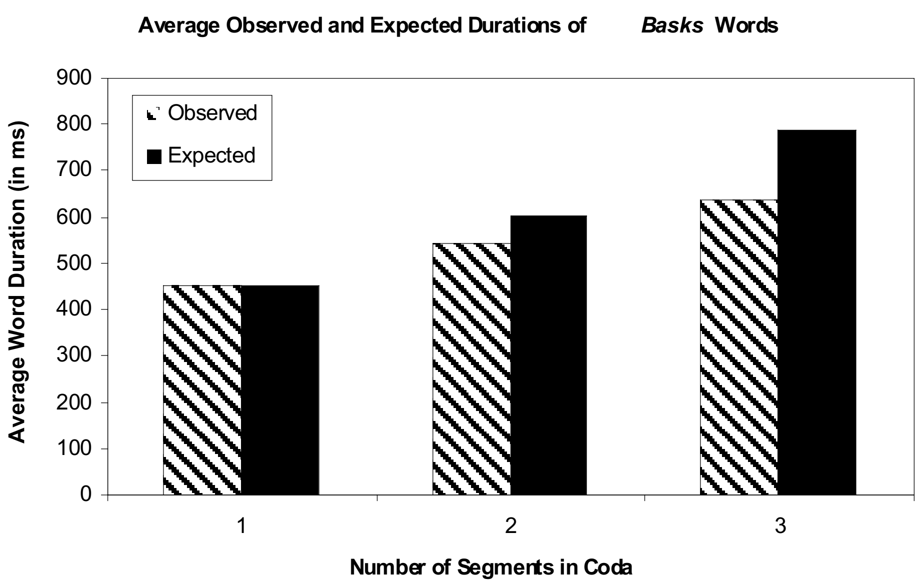

Figure 2. The above figure compares mean observed durations and mean expected durations of *basks* words based on the additive theory of word duration. The x-axis represents the number of segments in the coda and the y-axis represents the mean observed and expected durations of the words. This figure shows the same pattern found in

Figure 1. This pattern contradicts the Additive Theory but also refutes a strictly isocronal theory of overall word duration.

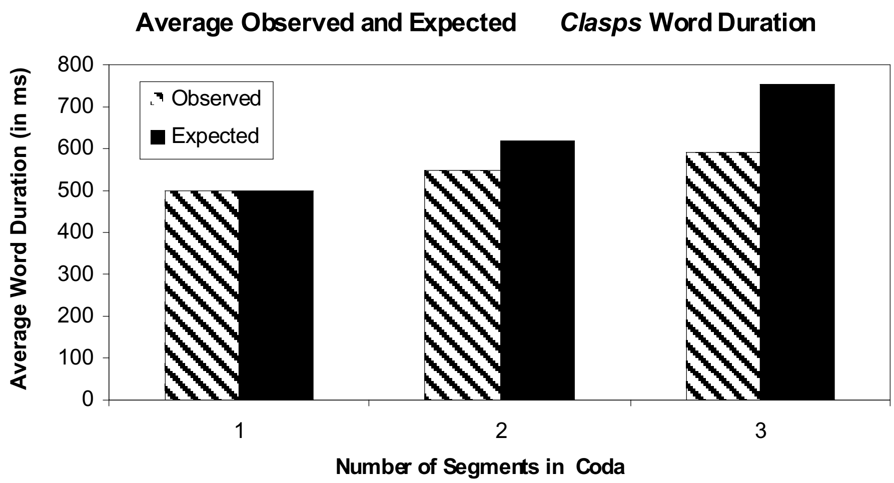

Figure 3. This figure compares mean observed durations with mean expected durations of *clasps* words in order to test the Additive Theory of Word Duration. The x-axis represents the number of segments in the coda and the y-axis indicates mean word duration in milliseconds. The pattern found here is the same as for Figures 1 and 2; all three provide evidence against the Additive Theory but also illustrate that these words are not perfectly isocronal either.

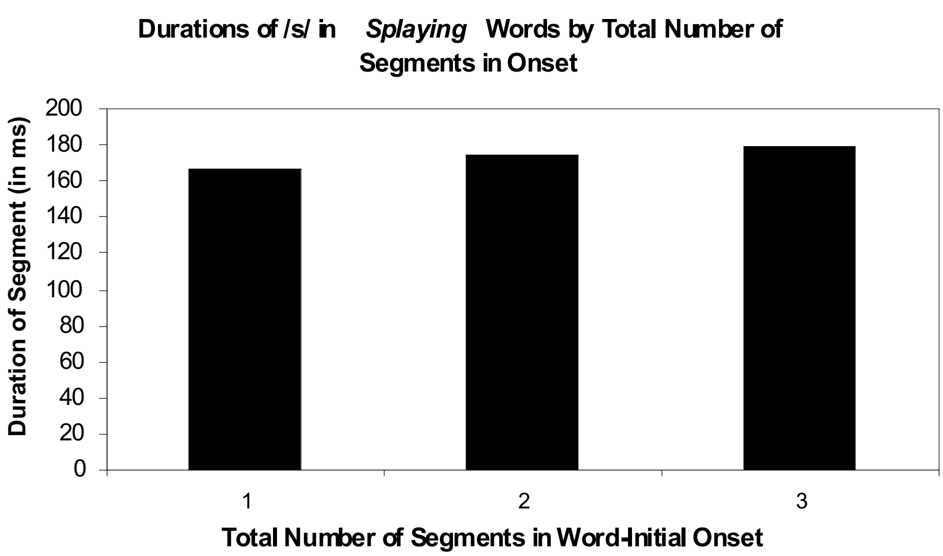

Figure 4. The above figure illustrates the mean durations of /s/ segments in /s/-initial *splaying* words (*saying, spaying, slaying, splaying*). The x-axis represents the total number of segments in the word-initial onset of the *splaying* word from which the segment was measured. The mean duration of /s/ in milliseconds (averaged across the three speakers) is indicated on the y-axis. Therefore, the first column represents the mean duration of /s/ in *saying*. The second column represents the mean of the mean durations of the /s/ segments in *spaying* and *slaying*. The third column stands for the mean duration of /s/ in *splaying*. As the graph illustrates, the duration of /s/ need not decrease as the number of other segments in the onset increases. This argues against the proposal that segments in a word compensate for the addition of a segment by uniformly decreasing in duration.

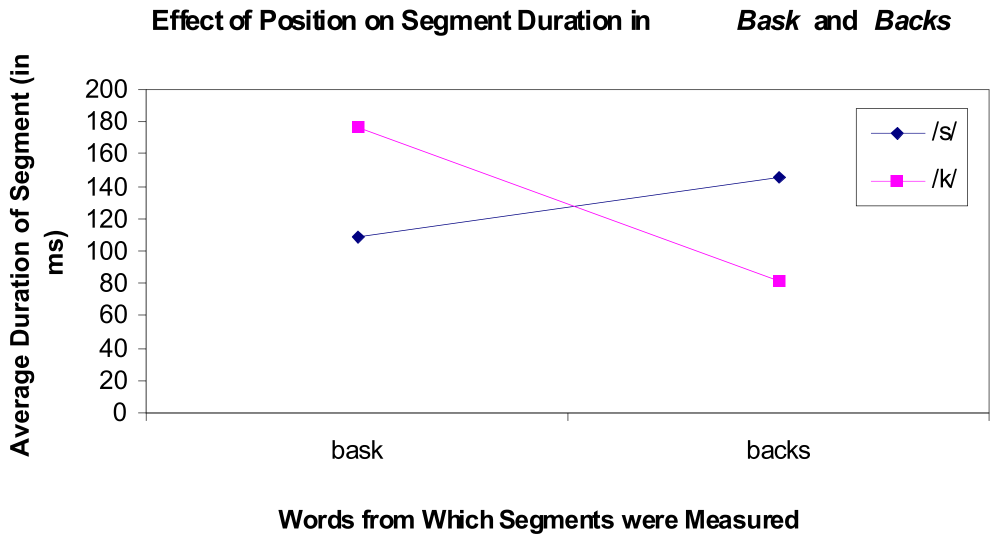

Figure 5. This figure illustrates the powerful effect of position on segment duration in consonant clusters. The x-axis indicates the word (*bask* or *backs*) from which the segments /s/ and /k/ were measured and the y-axis represents the mean durations of these segments in milliseconds, averaged across the three speakers. The pink symbols represent duration values for /k/ while the blue symbols represent duration values for /s/. If segment duration in a consonant cluster is determined by natural class, then we would expect for the durations of /k/ and /s/ to remain constant regardless of position. However, Figure 5 illustrates that this is clearly not the case. Rather, this figure provides evidence that segment duration in consonant clusters is highly influenced by position.

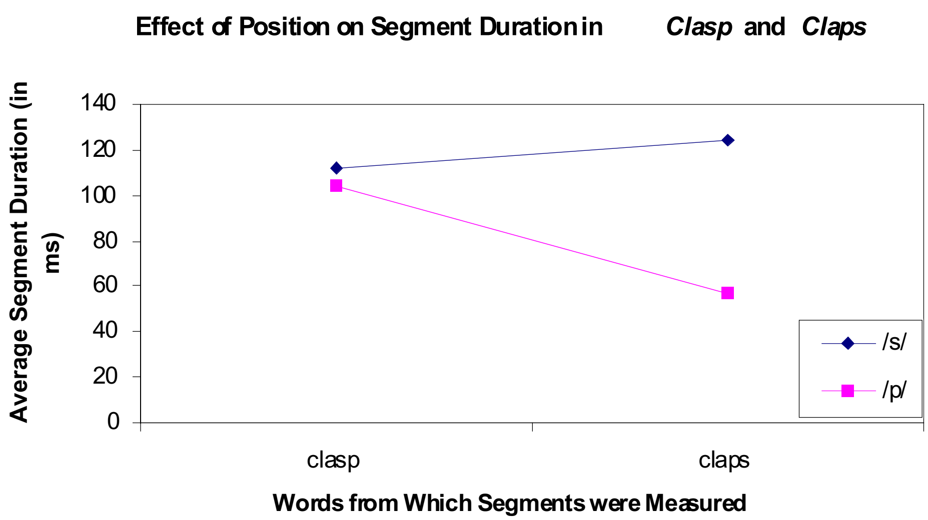

Figure 6. The figure shown above illustrates another example of the interaction between segment duration and segment position that was demonstrated in Figure 5. The x-axis indicates the words (*clasp* and *claps*) from which the segments were measured and the y-axis represents the mean segment duration in milliseconds. Here, the pink symbols represent the segment /p/ while the blue symbols again stand for the segment /s/.

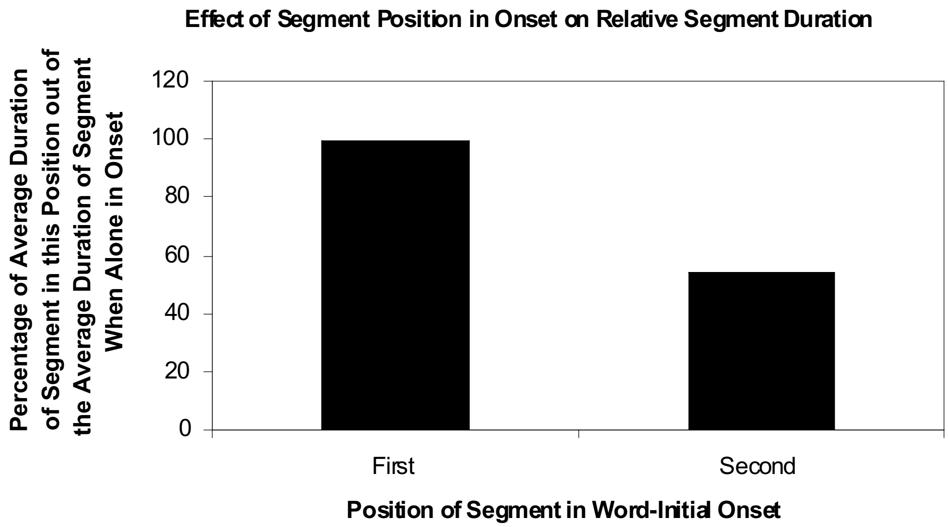

Figure 7. The above figure clarifies the effects of position on segment duration in onset consonant clusters. Specifically, these data come from measurements of the first and second segments (always either /s/ or /p/) in those *splaying* words that have two or three segments in their word-initial onsets. In order to average by position across segment type, the inherent duration of the segments must be taken into account. Therefore, the y-axis represents the percentage of the mean duration of segments in cluster contexts out of their mean duration when alone in an onset. For example, the duration of the /s/ in *spaying* was divided by the duration of the /s/ in *saying*. Once these values were determined for each segment in a cluster context, they were grouped by their position in the word and the mean percentage was calculated for each group. The x-axis indicates the two positions measured: first and second. The figure confirms that position determines the extent of segment duration reduction. Moreover, we can see the nature of this effect in onset consonant clusters; as segments are added, the first segment in the onset undergoes little to no duration reduction, while the duration of the second segment is substantially reduced.

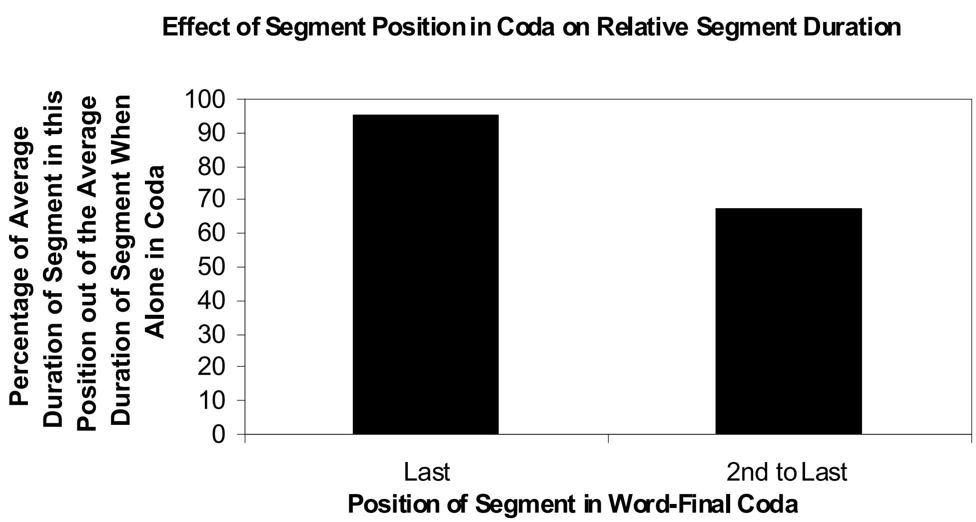

Figure 8. This figure clarifies the effects of position on segment duration in coda consonant clusters by representing the data from *basks* and *clasps* words. The x-axis indicates segment position within the word and the y-axis represents the mean duration percentage of the segments. (See Figure 7 for an explanation of the percentage values.) In Figure 8 we see that as new segments are added, the duration of the last segment undergoes little to no change while the duration of the second to last segment is drastically reduced.