

広島大学学術情報リポジトリ  
Hiroshima University Institutional Repository

Title	The effects of duration and frequency of occurrence of voiceless fricatives on listeners' perceptions of sound prolongations
Author(s)	Kawai, Norimune; Healey, E. Charles; Carrell, Thomas D
Citation	Journal of Communication Disorders , 45 (3) : 161 - 172
Issue Date	2012
DOI	<a href="https://doi.org/10.1016/j.jcomdis.2012.02.002">10.1016/j.jcomdis.2012.02.002</a>
Self DOI	
URL	<a href="https://ir.lib.hiroshima-u.ac.jp/00034814">https://ir.lib.hiroshima-u.ac.jp/00034814</a>
Right	(c) 2012 Elsevier Inc. All rights reserved.
Relation	



Manuscript Number: JCD-D-11-00036R2

Title: The Effects of Duration and Frequency of Occurrence of Voiceless Fricatives on Listeners' Perceptions of Sound Prolongations

Article Type: Research paper

Keywords: stuttering; sound prolongations, auditory perceptions

Corresponding Author: Dr. Norimune Kawai, Ph.D.

Corresponding Author's Institution: Hiroshima University

First Author: Norimune Kawai, Ph.D.

Order of Authors: Norimune Kawai, Ph.D.; E. Charles Healey, Ph.D.; Thomas D Carrell, Ph.D.

Abstract: The present study examined listeners' identification and rating of words in passages as "stuttered" when the duration and frequency of occurrence of sound prolongations were manipulated. Thirty-six participants listened to a 219-word passage containing voiceless fricatives digitally increased from their normal durations to 200, 300, and 420 ms. Listeners heard one of three passages that contained 5%, 10% or 15% altered stimuli within the passage. In Condition 1, listeners identified words considered "stuttered." In Condition 2, listeners rated specifically selected words in the passage relative to the extent they considered the words "stuttered." The results showed that 1) both the duration and the frequency of occurrence of the altered phonemes within the paragraph length material had an impact on listeners' perception of words identified as a sound prolongation; 2) listeners gave significantly higher ratings in Condition 2 than Condition 1 when determining if a word was stuttered or produced fluently. The implications of these results are discussed.

## COMMENTS AND SUGGESTIONS OFFERED BY REVIEWER 1

### Title and Abstract

The title wording remains fine. The authors have edited the abstract so that the description of the conditions is clearer.

The first sentence of the abstract needs editing for clarification. The "words" themselves are not considered to be "sound prolongations" (they are considered to have or contain sound prolongations). The authors could insert "to contain" to revise as: ".considered to contain sound prolongations." Or another revision improving the entire sentence might be: "The present study examined listeners' identification and rating of words in passages as "stuttered" when the duration and frequency of occurrence of sound prolongations on voiceless fricatives were manipulated."

• ***We agree with the suggested revision and have included the new wording in the abstract.***

### Introduction

Introduction, top of page numbered "1". The wording is hard to follow for the results of Schaeffer & Eichorn. More specifically, consistent wording is needed regarding whether listeners evaluated "phrases" or "vowels." In addition, readers might find it helpful to find information about how long a vowel becomes when it is lengthened by 150% and 250% that the authors considered to be only "slightly longer" (see authors' inference statement). Finally, because listeners judged only whether the spoken phrase sounded natural or not, the authors should make only a guarded inference about the perception of stuttering. Suggestions for possible editorial revisions: "The results compared to 83% who judged the phrases with vowels extended by 150% as natural." Revision to the next sentence: "Thus, making vowels within a word slightly longer (by about 300 msec) alters listeners' judgments of naturalness, and in turn, may impact their perceptions of whether a word is fluent or stuttered."

• ***All of these suggested changes have been made. Those changes can be found in the Introduction in the second paragraph.***

Introduction, middle of page numbered "1". Awkward wording: ".in 20 ms segments until 500 ms had been achieved." Clarification suggested: ".".

until a range of durations up to 500 ms had been achieved."

- ***Wording change were made in the third paragraph.***

Introduction, middle of page numbered "1". When the authors revised the sentence that starts "Based on methodology." they missed the concern with the last sentence of the previous paragraph that was raised by this reviewer. Note that this reviewer cannot quarrel with a statement of authors' conclusions (1st sentence of the last paragraph starting "Based on.") because the restatement of the authors' conclusions is a point of fact regardless of whether that conclusion happens to be an accurate or inaccurate one. But because categorical perception (of fluency vs. stuttering) was not specifically tested by the Kawai, Healey & Carrell (2005) study, this reviewer can argue that the last sentence of the previous paragraph describing the results should be made as a point of fact without the inference about continuous vs. categorical perception. To justify the argument that categorical perception was not tested: consider that just because a listener might be able to judge that there is a difference between a 5 ms and a 20 ms VOT along the VOT continuum does not address the question of whether /p/ vs. /b/ are judged categorically (also see this reviewer's later comments about the 1-100 rating scale in the review of the Discussion section). Suggested editorial revision: "The results showed that listeners perceived the gradual increases in fricative duration along the entire continuum of segment durations."

- ***We understand the reviewer's concerns and accept the change offered as shown above. Although it wasn't suggested, we thought it would be best to start a new paragraph with the "Based on the methodology...We hope this is acceptable.***

pp.4-5. The three research questions need to be worded so that they more closely match the analyses that were conducted:

Suggested revision of question 1): What percentage of the words with manipulated word-initial voiceless fricative durations are identified by listeners as "stuttered" for each of the three durations (200, 300, 420 ms) and three frequencies of occurrence (5%, 10%, 15%) within the connected speech passages?

[compare to the original and note that the percentage of words, not listeners, has been analyzed in this study]

Suggested revision of question 2): How do listeners rate the overall passage (along a continuum from fluent to stuttered) for the three durations of word-initial voiceless fricatives (200, 300, 420 ms) at

three frequencies of occurrence (5%, 10%,15%)?

[compare to the original and note that passages, rather than durations, are analyzed]

Suggested revision of question 3): When asked to focus specifically on each target word within the passage, how do listeners rate the words ( along a continuum from fluent to stuttered) based on three durations of word-initial voiceless fricatives (200, 300, 420 ms) and three frequencies of occurrence (5%, 10%,15%)? [compare to original and note that what was being rated (the continuum) was missing previously]

***• We have changed the wording as suggested for research questions 1 and 3 but in part, do not agree with the new wording for question 2. The second question did not ask listeners to rate the “overall passage.” They were asked to rate the speaker’s overall speech. This was also the instruction to listeners shown in the Method section 2.4, fourth paragraph. Thus, we used “speaker’s overall speech” rather than the reviewer’s recommendation of “overall passage.” The other suggested wording changes relative to question 2 were acceptable to us.***

## Methods

p.8. Section 2.3. Last paragraph. It would be better not to refer to the manipulated prolonged sound targets as "stuttered" words presumptively. Recommended revision: "Condition 2 was designed to examine how listeners perceived the manipulated words (i.e., with prolonged sounds)."

p.9, Section 2.4, last paragraph, last sentence. Minor edit: "The three forms. were (vs. was) presented."

p.11. Section 2.5. first paragraph. Because this first sentence appears to be the overview of the whole condition, it appears to want the added information to the effect that: "After the Condition 2 task was completed, listeners again rated the entire passage along the fluency scale, just as they did after completing the Condition 1 task."

***• We accept and have made all of these editorial changes. We would add that we made slight changes in wording beyond the reviewer’s recommendation. The new wording can be found in the last two sentences at the end of Section 2.5.***

## Results

p.15. 1st paragraph. The interaction findings appear to be similar, with parallel results across the three passages. It would be helpful

for this similarity to be highlighted for readers. Instead of "A second finding.(continue to the end of the paragraph)" consider replacing those two last sentences with something like: "Similar results were also obtained for the 10% passage and the 15% passage: ratings were significantly lower for the 200 ms durations relative to the 300 ms or 420 ms durations, but the ratings of 300 ms and 420 ms durations did not differ significantly."

**• We accept this change and have modified the manuscript to reflect the new wording in the last two sentences of Section 3.2.**

p.15 Section 3.3. Sentence 3. Should this sentence refer to Figure 2 rather than Figure 1? Otherwise, Figure 2 is not mentioned anywhere else in the body of the manuscript.

**• We appreciate the reviewer pointing out that we referred to the wrong figure in Section 3.3. The correct one is actually Figure 3, not Figure 2. We also have changed the wording of the labels in Figure 3 to reflect that the figure compares the overall ratings in Condition 1 with Condition 2. Thus, we have changed the labels from Pre-attentive and Focused Attention to Condition 1 and Condition 2, respectively. We also added Figure 2 to the second sentence under Section 3.2.**

pp.15-16 in Section 3.3 regarding Overall Speech Ratings. The analysis of differences between Conditions 1 and 2 is interesting, but it seemed like obtaining overall ratings would have led to an analysis of whether duration vs.frequency of occurrence had the greater impact on overall ratings of the passage. This is to me, an interesting question that I wished the authors had addressed in their analysis of the Overall Speech ratings: Which of the two variables, duration vs. frequency of occurrence, has the greater impact on the perception of the passage as stuttered? There is an impression based on the results (see either Table 3 or 4) that in contexts like these data (where duration of the fricative sound prolongation is under 500 ms), that the frequency of occurrence (not duration of the sound prolongation) has the stronger impact on the perception of the connected speech as "stuttered", even though listeners are readily able to detect the durational differences among the prolonged sounds from 200ms to 420 ms.

**• The reviewer brings up an interesting point about whether duration or frequency had the greater impact on listeners' overall ratings. Although the reviewer suggests that frequency has the stronger influence, we do not agree that our data show that. The purpose of Section 3.3 was simply to show the difference in ratings between conditions. Our statistical results show an interaction effect, which means that duration and frequency in combination affected the ratings.**

***Therefore, we do not believe that our data cannot or should not address the question or issue the reviewer raises.***

## Discussion

Section 4.1. parag. 1. sent. 1. The wording of each of the first two sentences needs revision, because "increases in duration" (bottom of p. 16) do not result from the identification and rating procedures, rather it is the other way around. Also, words are not perceived as sound prolongations (top of p.17), but as having or containing sound prolongations.

***• We have made a change to the first sentence of the Discussion section (4.1). We trust the new wording is acceptable.***

Although in the authors response to the review, it is indicated they intended to address the matter in section 4.2, the authors still have not acknowledged the potential issue of perceptual bias demonstrated by Williams & Kent (1958). This early study showed that the label listeners are instructed to apply affects the frequency of identifying that target.

So in the present study, the authors need to consider how results might have been different if the listeners had applied different labels. For example, what if the listeners had been given the opportunity to apply the label "nonfluent, but not stuttered" in addition to the two labels of "fluent" vs. "stuttered"? We do not know whether the percentage of judgments of "stuttered" might have changed if another label had been offered as an alternative. See again: Williams, D., & Kent, L. (1958). Listener evaluations of speech interruptions. *Journal of Speech and Hearing Research*, 1, 124-131.

Although ratings on a scale from 1=fluent to 100=stuttered (with ratings of 2-99 representing the range of uncertainty inbetween) might appear to have a straightforward interpretation, I am not convinced that the interpretation of these data are straightforward and handled appropriately in this Discussion. Some listeners could conceive of the range as being from 1= most fluent to 99=least fluent, but "stuttered" only at 100, while other listeners could consider 1=fluent and 2 as disfluent/possibly stuttered to 100 severely/unambiguously stuttered. So in fact, the point at which a listener judges the duration as "stuttered" along the continuous rating scale from 1=fluent to 100=stuttered, is not really known. It is clear that listeners detect the gradients in the durations of sound prolongations along the continuum from shorter to longer based on their ratings along the scale, but their concept of the label "fluent" or "stuttered" along that continuum cannot be inferred

conclusively. In fact, according to the former 100=stuttered concept, it could be argued that none of the words in Table 2 were judged to be fully and unambiguously "stuttered" because all the mean ratings were less than 100. I think the authors would do well to acknowledge and address this limitation of this scaling procedure in their Discussion section.

***• We have addressed these issues more completely by adding a new paragraph to the end of Section 4.1 (see the last paragraph). We trust this addresses this reviewer's concern more completely. If not, we are willing to modify it as necessary.***

pp.19-20. I think that the authors could consider whether these results have an implication important for speech clinicians working with families of people who stutter. The overall rating differences between conditions 1 and 2 suggest that if a listener focuses on whether selected words are stuttered within a connected speech sample, that it could lead to a greater impression of stutter-like speech than when the listener is asked to merely to listen for whether there are any words in a connected speech sample that are stuttered.

***• We are a little reluctant to comment on the clinical implications of our findings because the study focused only on sound prolongations and any interpretation of our results to other forms of nonfluent speech would seem inappropriate. However, we did attempt to address the reviewer's suggestion by adding a short paragraph to the end of Section 4.2. In some respects, we think that what we have written might be a bit of a stretch but are willing to modify or delete it. Feedback on what we have written would be helpful.***

p.20. 1st paragraph, 2nd complete sentence. Editorial revision and clarification are needed for the sentence starting: "In the passage containing 15% altered stimuli, listeners identified more frequency moments of stuttering." Are the authors trying to describe Condition 1 or 2? I think this is supposed to be referring to Condition 1. So perhaps wording would be clearer as: "Comparing results across passages in Condition 1, listeners identified a higher frequency of the target words as stuttered in the passage with 15% altered stimuli compared to 5% altered stimuli. Thus, as listeners."

***• In the second to last paragraph in Section 4.2, we have clarified that we were referring to Condition 2. We think that is clearer now than it was. We then modified the next sentence and deleted the sentence "Thus, as listeners were exposed to more moments of sound prolongations, they may have increased their attention toward moments of sound prolongations." We then kept the next two sentences from the previous version of the manuscript.***



p.20 Section 4.3. 1st sentence. Minor edit:. "Although this the first study" should be "Although this is the first study."

- ***Changed as suggested***

#### Tables & Figures

The tables and figures are helpful to the interpretation of results. Be sure to mention Fig. 2 in the body of the manuscript.

- ***We added Figure 2 to the text and mentioned this above in an earlier comment.***

#### Other

I am not sure how the sentences preceded by > are supposed to function within the manuscript.

- ***We were not sure what the reviewer meant by this comment. We could not find any sentences preceded by > so if the reviewer to guide us to the exact location of these sentences, we can address this comment.***

There appear to be two sets of Appendix B. Continuing Education questions, but they appear to be the same. The questions and options are easier to understand than those submitted originally.

- ***We must have inadvertently kept the original set of questions in the manuscript rather than deleting it when we modified the CE questions. We apologize for the confusion. There is now only one Appendix B.***

For CE question 3., shouldn't the correct answer be (b) rather than (a)?  
I referred to results for Condition 1 reported in the first paragraph on p.14.

- ***The reviewer is correct. Question 3 should be (b), not (a). We also changed the answer key.***

The Effects of Duration and Frequency of Occurrence of Voiceless Fricatives on  
Listeners' Perceptions of Sound Prolongations

Norimune KAWAI<sup>a)</sup>, E. Charles HEALEY<sup>b)</sup> and Thomas D. CARRELL<sup>c)</sup>

<sup>a)</sup> The Center for Special Needs Education Research and Practice

Hiroshima University

1-1-1 Kagamiyama C708

Higashi-Hiroshima, Hiroshima 739-8524 JAPAN

TEL & FAX: +81-82-424-7179

Electronic mail: nkawai@hiroshima-u.ac.jp

<sup>b)</sup> Department of Special Education and Communication Disorders

University of Nebraska-Lincoln

353 Barkley Memorial Center

Lincoln, NE 68583-0731 USA

Electronic mail: ehealey1@unl.edu

<sup>c)</sup> Department of Special Education and Communication Disorders

University of Nebraska-Lincoln

318Q Barkley Memorial Center

Lincoln, NE 68583-0731 USA

Electronic mail: tcarrell@unl.edu

Running title

Listeners' Perceptions of Sound Prolongations

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

The Effects of Duration and Frequency of Occurrence of Voiceless Fricatives on Listeners’  
Perceptions of Sound Prolongations

# Abstract

The present study examined listeners' identification and rating of words in passages as "stuttered" when the duration and frequency of occurrence of sound prolongations were manipulated. Thirty-six participants listened to a 219-word passage containing voiceless fricatives digitally increased from their normal durations to 200, 300, and 420 ms. Listeners heard one of three passages that contained 5%, 10% or 15% altered stimuli within the passage. In Condition 1, listeners identified words considered "stuttered." In Condition 2, listeners rated specifically selected words in the passage relative to the extent they considered the words "stuttered." The results showed that 1) both the duration and the frequency of occurrence of the altered phonemes within the paragraph length material had an impact on listeners' perception of words identified as a sound prolongation; 2) listeners gave significantly higher ratings in Condition 2 than Condition 1 when determining if a word was stuttered or produced fluently. The implications of these results are discussed.

# The Effects of Duration and Frequency of Occurrence of Voiceless Fricatives on Listeners' Perceptions of Sound Prolongations

## 1. Introduction

In order to understand what listeners consider stuttering versus normally fluent speech, it is important to determine specific temporal features of words and sounds that listeners identify or consider as a disruption in the forward flow of speech. Research on listeners' perceptual labels of normally fluent and stuttered speech began several decades ago. Sander (1963) and Curran and Hood (1977) demonstrated that listeners' perceptions could be manipulated through small changes in the speech sample that would create a shift in the perception of words from fluent to stuttered. Specifically, Sander changed part-word repetitions from one- to two-units per repetition (e.g, Sa-Saturday to Sa-Sa-Saturday) and found that it took only half the instances of double unit than single unit repetitions in the speech sample for listeners to judge the speech sample as stuttered. In a similar study, Curran and Hood modified whole-word repetitions by having the word repeated twice instead of once. More recently, Amir and Yairi (2002) discovered that listeners considered short interval and vowel durations of repeated units as stuttering-like while vowel and interval durations that were digitally prolonged were judged as more normal-like productions.

Given that repeated sounds and words could be experimentally manipulated, is it possible to change a consonant or vowel within a word from fluent to one that is perceived to be a sound prolongation? In an effort to answer this question, Schaeffer and Eichorn (2001) studied how listeners perceived gradual increases in the duration of vowels embedded in a carrier phrase (i.e., "say \_\_\_\_ again"). Instead of investigating the point at which listeners consider sounds in a word as "normal" or "abnormal" in duration, they increased the vowel duration of target words by

150%, 200%, and 250%. The authors sought to determine which percentage of vowel increase influenced listeners' perception of speech naturalness. The results showed that only 22% of the listeners judged the phrase with vowels extended 250% as natural compared to 83% who judged the phrase with vowels with 150% extension as natural. Thus, making vowels within a word slightly longer (by about 300 ms) alters listeners' judgments of naturalness, and in turn, may impact their perceptions of whether a word is fluent or stuttered.

Although Schaeffer and Eichorn (2001) manipulated sound durations to determine shifts in listener perceptions of speech naturalness, would a similar perceptual change occur for the labels of "fluent" vs. "stuttered" words if the initial sounds were systematically increased across a wide range of durations? In an attempt to answer this question, Kawai, Healey, and Carrell (2005) digitally manipulated the durational boundaries of a word contained in an isolated sentence. The natural 120 ms duration of a voiceless fricative was varied in 20 ms segments until a range of durations up to 500 ms had been achieved. Listeners were instructed to rate each initial-word fricative segment on a scale from 1-100 (1 = fluent, 100 = stuttered). The results showed that listeners perceived gradual increases in the fricative duration along the entire continuum of segment durations.

Based on the methodology used to interpret listeners' responses, Kawai, Healey, and Carrell concluded that listeners do not perceive a specific perceptual boundary for words considered "fluent" or "stuttered." A range of durations of the initial voiceless fricative created several levels of uncertainties of when the target word was perceived as a sound prolongation. Nonetheless, the data from Kawai, Healey, and Carrell (2007) showed that initial-word durations ranging from 120 to 200 ms were perceived as fluent" and the durations in the range of 420-500 ms were perceived as stuttered. Zebrowski (1991, 1994) also found that listeners perceived

1  
2  
3  
4 durations of sound prolongations longer than 435 ms as stuttering. Initial phoneme durations  
5  
6 between 200 and 400 ms (i.e., ~300 ms) would be labeled somewhat ambiguously or “abnormal”  
7  
8 in a study by Lingwall and Bergstrand (1979) because this duration would occur in between the  
9  
10 range of what most listeners perceive as fluent and stuttered words.  
11  
12  
13

14 A potential limitation of past studies in this area is that all have used limited linguistic  
15  
16 contexts from which to obtain perceptual judgments from listeners. For example, in the Kawai et  
17  
18 al. (2007) study listeners heard each target word within an isolated phrase extracted from  
19  
20 paragraph material. Jones, Logan, and Shrivastav (2005) embedded a target word within a  
21  
22 sentence repetition task while Schaeffer and Eichorn (2001) used a carrier phrase. The brief  
23  
24 speech samples from these previous studies provide experimental control over the stimuli but  
25  
26 listeners have limited perceptual information from which to rate or label a word as fluent or  
27  
28 stuttered. An expansion of our scientific knowledge about listeners' identification of sound  
29  
30 prolongation changes could be obtained by including various voiceless fricative durations in  
31  
32 words embedded in paragraph-length speech material in which a short story is heard. In typical  
33  
34 conversational speech, listeners hear a speaker produce a series of utterances rather than short,  
35  
36 isolated utterances. Thus, it would be more realistic if the presence of sound prolongations were  
37  
38 judged within a connected speech context rather than from isolated sentences. One of the aims of  
39  
40 the present study was to determine how listeners perceive sound prolongations while listening to  
41  
42 a short story. Perhaps listener perceptions of stuttering are altered when listening to a story rather  
43  
44 than short sentences typically used in previous studies.  
45  
46  
47  
48  
49  
50  
51

52 Another factor that could influence whether listeners' judgments of speech samples as  
53  
54 stuttered or fluent is the number of words that are modified within the sample. Previous research  
55  
56 by Hegde and Hartman (1979) found that whole-word repetitions repeated once are regarded as  
57  
58  
59  
60  
61  
62  
63  
64  
65

normally fluent but when the frequency of those single unit repetitions were presented on 15% of the words in a 100-word paragraph, the majority of listeners reported they heard stuttering. Susca and Healey (2001) discovered similar results. They found that when the number of stuttered words was increased from 5% to 15% within paragraph-length material, listeners gave increasingly negative ratings of the speaker's fluency and communicative competence.

In the Hegde and Hartman (1979) and Susca and Healey (2001) experiments, listeners rated the speech sample as being representative of fluent or stuttered speech. Neither of these studies asked listeners to identify specific words that were considered stuttered that led listeners to report a particular perceptual rating. The importance of investigating how well listeners identify specific words as "stuttered" within paragraph-length material will add to our understanding of how this context might affect listeners' durational thresholds when identifying a sound as stuttered. Moreover, having listeners focus on a specific word to judge whether it is prolonged or not assures that listeners are attending to targeted temporal information. To date, no research has compared listeners' perceptions of different types of stuttering when they do and do not focus on specific sounds or words within a speech sample. One could speculate that adding or deleting word disruptions and calling a listener's attention to a specific word in a speech sample might produce a different rating than if listeners were asked to judge if any stuttering occurred in an entire sample of speech. In natural communicative situations, listeners typically don't focus on just sounds or words. Rather, they focus on several aspects of how the person's speech is produced (Susca & Healey, 2001).

From a theoretical perspective, Feature Integration theory (Treisman, 1986; Treisman & Gelade, 1980; Treisman & Schmidt, 1982) proposes a two-stage perceptual system that determines the existence of the basic units of perception. In the pre-attentive stage, the perceptual



system analyzes a perceptual image and determines the existence of the features. In the focused attention stage, signal features are combined to create a perceptual image, which is then compared to the perception of information stored in memory. This theory could serve as a foundation for studying differences in listeners' perceptions of focused and unfocused attention to words within a passage, would add to our understanding of what prompts listeners to perceive a word as stuttered, fluent, or somewhere in between.

In summary, previous research has demonstrated that listeners' identification of words as fluent vs. stuttered can be influenced by how many times a sound or word is repeated or the durational increase of a phoneme that turns a fluent sounding word into one that resembles a sound prolongation. However, it has yet to be determined if the perception of initial-word sound durations (i.e., digitally manipulating ones that are considered fluent to ones that are considered stuttered) can be altered by changing the frequency of their occurrence within a paragraph-length speech sample. In past studies, researchers have measured listeners' overall reactions to a speech sample containing various types and frequencies of stuttering. No research has been conducted to determine how speech is perceived when listeners are asked to identify specific words within a speech sample that they believe are "stuttered."

The purpose of the current study was two fold: 1) to examine listeners' identification and rating of words that are considered stuttered when the duration and frequency of occurrence of voiceless fricatives were manipulated within paragraph-length speech material, and 2) to examine listeners' ratings when listeners were asked to rate the duration and frequency of occurrence of specifically selected word-initial voiceless fricatives within paragraph-length speech material. Specifically, three research questions were addressed:

1) What percentage of words with manipulated word-initial voiceless fricatives are identified by listeners as stuttered for each of the three durations (200, 300, 420 ms) and three frequencies of occurrence (5%, 10%, 15%) within the connected speech passages?

2) How do listeners rate a speaker's overall speech (along a continuum from fluent to stuttered) when word-initial voiceless fricatives are manipulated at three durations (200, 300, 420 ms) at three frequencies of occurrence (5%, 10%, 15%) within the connected speech passages?

3) When asked to focus specifically on each target word with a passage, how do listeners rate the words (along a continuum from fluent to stuttered) based on three durations of word-initial voiceless fricatives (200, 300, 420 ms) at three different frequencies of occurrence (5%, 10%, 15%) within the connected speech passages?

## 2. Method

### 2.1 Participants

A total of 36 adult listeners (18 males and 18 females) participated in the study. The males ranged in age from 19 to 43 years ( $M = 23.94$ ,  $SD = 6.36$ ) while the females ranged in age from 19 to 40 years ( $M = 23.61$ ,  $SD = 6.71$ ). None of the listeners were graduate students majoring in speech-language pathology, certified speech-language pathologists, faculty in speech-language pathology, family members of people who stutter, or anyone who had extended contact with people who stutter (i.e., a co-worker or close friend). The intent was to find naïve listeners who had minimal exposure to stuttering behavior prior to participating in this study. Thus, the listeners collected for this study would most likely reflect a typical listener's response to what is considered fluent and stuttered speech. All listeners' spoke English as their native language and none reported having any history of speech, language, or neurological disorders.

To verify that all participants had normal hearing, each participant had to pass a standard hearing screening consisting of a range of pure tones (i.e., 500 to 8K Hz) presented at 20 dBHL.

## 2.2 Stimuli and Experimental Passages

A modified version of a three paragraph, 219-word passage about fossils (Murphy, 1989) was used as the experimental stimulus material because it contained a large number of initial-word voiceless fricatives. The passage was modified from its original version by: (1) increasing the number of word-initial voiceless fricatives within the passage in order to create a high frequency of words initiated with voiceless fricatives that listeners might be able to perceive as stuttered, and (2) adding words and sentences between target sounds so that a minimum two seconds elapsed between manipulated words to allow for listeners to make a perceptual judgment (Appendix A).

Voiceless fricatives were chosen as the target experimental stimuli because: (1) it is easy to digitally extend the duration of the frication noise using a speech analysis computer program (Sound Forge ver. 4.0c, 1997), (2) the normal duration of all voiceless fricatives are approximately the same (Klatt, 1973; Umeda, 1977), and (3) word-initial voiceless fricatives could be produced as sound prolongations by PWS. Of the 47 word-initial voiceless fricatives within the passage, 8 were not digitally modified because the words contained a consonant cluster (e.g., “from” and “storm”) or were adjacent to other voiceless fricatives. Thus, a total of 39 voiceless fricatives served as target stimuli.

The original duration of each voiceless fricative ranged from 98-147 ms ( $M = 120.49$ ,  $SD = 14.53$ ) but the duration of all 39 voiceless fricatives within the passage were digitally manipulated resulting in a mean duration of 120 ms, which was within normal limits for all the word-initial voiceless fricatives within the passage. Next, the first author created three separate

experimental passages in which 5%, 10%, or 15% of the total words had extended voiceless fricative durations in order to represent mild, moderate, and severe stuttered speech as defined in a previous study by Susca & Healey (2001). A total of 11 voiceless fricatives within the 219-word passage were extended in order to create a passage with 5% prolonged voiceless fricatives. For the passage that contained 10% prolonged voiceless fricatives, the durations of 22 word-initial fricatives were digitally manipulated. The durations of 33 word-initial voiceless fricatives were digitally manipulated to create a passage with 15% prolonged sounds as target words. The remaining six voiceless fricatives were not manipulated and presented at normal, fluent durations in all three passages to prevent listeners from being able to detect that all voiceless fricatives within each passage were modified. The locations of these unaltered voiceless fricatives were at the beginning, middle, and end of the passage.

Within the three passages, targeted word-initial voiceless fricatives were arbitrarily increased from 120 ms. to 200, 300, and 420 ms. durations. A 200 ms duration extension was chosen because it represented a slight increase of a typical word-initial voiceless fricative. Extending the voiceless fricative from 120 to 300 ms represented a sound duration that was positioned between the fluent and stuttered perceptual continuum, as determined from the Kawai et al. (2007) study. That study also showed the majority of listeners perceived word-initial voiceless fricatives as stuttered when the duration of a sound prolongation was 420 ms or longer, which is why that duration was also used in the present study.

The altered word-initial voiceless fricatives within the three passages were distributed evenly across the beginning, middle, and end of any one passage. Because listeners were required to make many judgments throughout the three passages, the target words chosen for manipulation were different for each passage and a minimum of two seconds was created

between any two manipulated sound/word durations. All voiceless fricative durations were digitally manipulated via a speech acoustic analysis computer program (Sound Forge version 4.0c, 1997).

Prior to the manipulation of the word-initial voiceless fricatives, it was necessary to create a recording of the experimental passage. A 23-year-old male graduate student at the University of Nebraska-Lincoln who had no speech, language, hearing, and/or neurological disorders read the experimental passage. The speaker's reading rate was 196 syllables per minute, which was within normal limits (Klatt, 1973; Umeda, 1977; Walker, 1988).

Listeners were exposed to only one frequency-altered passage (5%, 10%, or 15%) because individuals who listened to the 15% sample first would have been biased if they listened to the 5% sample later in the experiment. Therefore, each participant listened to only one of the three frequency-altered passages. For example, if a participant was assigned the 5% passage, that person heard one presentation with 200 ms altered stimuli, another one with 300 ms, and another with 420 ms altered stimuli. A total of 12 participants (6 males and 6 females) were tested at each target frequency (5, 10, or 15%) and exposed to each of the three durations of altered stimuli (200, 300, 420 ms). Distribution of ages across the three target frequencies was relatively similar.

### 2.3 Experimental Conditions

The experiment was divided into two conditions. Condition 1 examined the percentage of words listeners perceived as stuttered when they listened to the three experimental passages containing acoustically altered voiceless fricatives. The purpose of this condition was to simulate a real listening condition in which listeners would hear words that were fluent and prolonged at different durations that might be perceived as stuttered. A secondary part of this condition was to

have a listener rate the overall speech of the individual reading the passage on a scale of 1 (fluent) to 100 (stuttered).

Condition 2 was designed to examine how listeners perceived manipulated words (i.e. with sound prolongations) when they were directed to focus on specifically selected target words in each of the same three experimental passages heard in Condition 1. The purpose of this condition was to examine how listeners rated specific words containing extended durations on a scale of 1-100 and as in Condition 1, provide an overall rating of the speech on a scale of 1 (fluent) to 100 (stuttered).

Listeners participated in both experimental conditions. The order of conditions was not counterbalanced because all participants had to complete Condition 1 first to allow them to identify and rate on a scale of 1-100 any words they considered stuttered. They were not told which words involved some manipulation of the voiceless fricative.

Both conditions were conducted in a sound treated room. Prior to the experiment, the ambient noise of the room was at an acceptable level (approximately  $31.0 \pm 3$  dB SPL).

## 2.4 Procedures for Condition 1

Once listeners passed the hearing screening, they were seated at a booth in the room where the listening experiment took place. A 17-inch TFT-LCD computer monitor (Samsung Syncmaster 731B), connected with a 1.66 GHz Intel Mac Core Duo T2300 Macintosh computer (Apple Mac Mini Model A1176) with 1.96 GB of RAM and a Windows XP Pro 2002 Service Pack 2 platform was set up in the booth. The distance between listeners face and the display was approximately 60 cm. Then Beyerdynamic DT 770, circumaural, closed headphones were given to each participant so he/she would hear the presentation of the speech samples at a comfortable

loudness level (approximately 74 dB SPL). The average ambient noise emitted from the microphone was approximately 10.6 dB SPL or lower in both right and left sides.

Next, each listener received a printed copy of the experimental passage on plain white paper with 12-point font. Participants were told they would hear three versions of a passage about fossils. The three forms of the Fossil passage were presented separately and repeated once so listeners could verify the stuttered words they identified before going on to the next passage.

All auditory stimuli were played using the software Maketape version 2.2 (Srinivasan & Carrell, 2004). As participants heard each passage, they were instructed to place a check mark above any word they considered "stuttered." Once listeners heard each passage, they were told they would hear the passage a second time in order to give them a chance to verify their identification of stuttered words.

After each participant completed the identification task within Condition 1, they were instructed to look at a computer-based magnitude scale slider bar (The Pro Slider version. 2.0, Kawai, 2007) that appeared on the computer screen. At the top of the slider bar it said, "This person's overall speech is..." which was a prompt for the listener to rate the speech on a scale of 1-100. On the left side of the slider bar, there was a small window that showed a number (1-100) so that listeners were able to see the number they assigned to each speech sample. If listeners perceived the speaker's overall speech as stuttered, they were supposed to place the arrow using the computer mouse at "100" and if it was perceived as normally fluent, they were to place the arrow on the slider bar at "1." If they were unsure, they were instructed to place the arrow somewhere between "1" and "100." Listeners were not provided with any labels or definitions of perceptual categories between 1 and 100 and they were not allowed any practice in making ratings on the slider bar prior to or during the experiment.

The procedure described above was repeated each time for the next two passages that contained altered stimuli at one frequency of occurrence. A new blank copy of each new experimental passage was given to participants so they could check the words they perceived were stuttered. The order of the presentation of the passages containing 200, 300, and 420 ms altered durations at 5%, 10% or 15% occurrence was randomized across participants.

## 2.5 Procedures for Condition 2

Once Condition 1 was completed, participants began Condition 2. The procedures were similar to Condition 1 in that participants heard the three durations of altered words in the experimental passage at only one frequency of occurrence. However, the task for this condition was to have participants rate each altered word selected by the experimenters for each passage. Participants were told they would rate each word presented to them using the same slider bars and scales used in Condition 1. After that task, listeners rated the entire passage along the fluency scale, just as they did in Condition 1.

At the beginning of Condition 2, each participant was given another copy of the experimental passage. Before the passage was presented, a series of slider bars appeared on the computer screen, one for each target word from each of the experimentally altered passages. Each word appeared at the top of each slider bar relative to the number of words for the percentage of occurrence in each passage (i.e., 11 for 5%, 22 for 10%, and 33 for 15%). Participants were told that they would have to rate the word during the presentation of the passage on a scale of 1-100, which was the same scale used in Condition 1. They were also told they would have to complete the rating within two seconds. Participants were given two practice sessions prior to this experimental task for making ratings within two seconds of the presentation of the stimuli. All were successful in making ratings within this time frame.



Condition 2 began with a random presentation of one of the three experimental passages. Participants made their ratings while listening to the passage. Participants were then told to listen to the same passage a second time to allow them to verify their ratings that appeared at the top of the slider bars. Immediately following the verification of their ratings for each passage, listeners were asked to rate the overall speech of the speaker they heard. This involved the same type of rating made at the end of Condition 1.

## 2.6 Data Analyses

The independent variables for this study consisted of three voiceless fricative durations (i.e., 200, 300, and 420 ms) and three different frequencies of occurrence in the experimental passage (i.e., 5, 10, and 15%). The dependent variables were listeners' percentage of stuttered words identified in Condition 1. This was defined as the number of words listeners identified of the 11 target words that were manipulated in the 5% passage, the 22 target words in the 10% passage and the 33 target words in the 15% passage. The number of manipulated words each listener identified was divided by the total modified for the 5%, 10%, or 15% passages.

Listeners' ratings of the overall speech of the speaker were based on the rating given using a scale of 1 (fluent) to 100 (stuttered). In Condition 2, listeners were asked to rate specifically selected words that were identified for them by the researchers, which were the same number as above in Condition 1. They also rated the speaker's overall speech using the same 1-100 scale.

Means and standard deviations of identified stuttered words for all listeners were calculated. In addition, the standard deviation/mean ratio was calculated to provide a description of the degree of variability of listeners' ratings across the 5, 10, and 15% passages. A two-way mixed groups factorial ANOVA was used to determine how those effects influenced listeners' identification and ratings of digitally manipulated voiceless fricatives across the different sound

1  
2  
3  
4 durations (within-group factor) and frequencies of occurrence (between group factor) for each of  
5  
6 within three experimental passages. Effect sizes were reported as Eta-squared for all significant  
7  
8 findings.  
9

### 10 11 **3. Results**

#### 12 13 **3.1 Identification of Stuttered Words in Condition 1**

14  
15 In this condition, listeners were instructed to identify any of the 219 words they  
16  
17 considered stuttered (i.e., representative of a sound prolongation) while listening to the  
18  
19 experimental passages. Figure 1 shows participants' average percentage of identified stuttered  
20  
21 words associated with three altered initial voiceless fricative durations across the 5%, 10%, and  
22  
23 15% passages. In general, the data show that 64% of the listeners judged words as stuttered when  
24  
25 the initial voiceless fricatives were extended to 200 ms. These initial phoneme durations were  
26  
27 close to the fluent sound of the word (i.e. 120 ms) but substantially more participants identified  
28  
29 stuttered words at a high percentage rate when the initial voiceless fricatives were extended to  
30  
31 300 and 420 ms, 95% and 98% respectively. This finding held true for the 5%, 10%, and 15%  
32  
33 passages. The standard deviation (SD)/mean (M) ratios of listener identified stuttered words also  
34  
35 show that the largest variability occurred on words with 200 ms initial phoneme durations. Small  
36  
37 variability scores were found for initial voiceless fricatives with 300 and 420 ms durations at all  
38  
39 three frequencies of occurrence when compared to the variability at 200 ms durations. See Table  
40  
41 1.  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 To determine if any of the average percentage of identified stuttered words was  
52  
53 statistically significant, a 3 X 3 (Duration x Frequency) mixed groups factorial ANOVA with  
54  
55 follow-up analyses was used. The alpha level was adjusted to .045 via the false discovery rate  
56  
57 control method (Benjamini & Hockberg, 1995) to prevent a Type 1 error and was used to  
58  
59  
60  
61  
62  
63  
64  
65

examine the effects of the duration of the stimuli across the 5%, 10%, and 15% passages for listeners' identification of stuttered words beginning with a voiceless fricative.

Because the results showed there was a significant interaction between Duration and Frequency for the average percentage of identified stuttered words ( $F(4, 66) = 2.902, p = .028$ ), the main effects results will not be reported. The effect size was obtained from the statistical results summary showing the partial Eta-squared was .150. According to Cohen's (1988, 1992) rules (i.e., small = .1; medium = .3; large = .5), the effect size for this statistically significant finding was small. The simple effects showed that the altered stimuli in the 5% and 15% passages, for listeners' average percentage of identified stuttered words was significantly smaller for the 200 ms than for 300 ms and 420 ms durations. Listeners' average of identified stuttered sounds at 300 ms and 420 ms were not significantly different. Also, in the 10% passage, the listeners' average percentages of identifying the three altered phonemes were not statistically significant.

### 3.2 Ratings of Stuttered Words in Condition 2

In Condition 2, participants were asked to rate specific words selected for them within the experimental passages on a scale of 1-100 (fluent to stuttered). The average ratings for the three altered durations across the three frequencies of occurrence are shown in Table 2 and Figure 2. As found in Condition 1, the mean ratings were higher as the durations of altered voiceless fricatives became longer and in the 5% to 15% passages. Moreover, the variability as seen in the SD/M ratios was highest for the 200 ms duration.

In order to determine if any of the descriptive data were significantly different, a 3 X 3 (Duration X Frequency) mixed group factorial ANOVA was conducted using an adjusted alpha level of .045 to test the main and interaction effects of listeners' ratings of specific sounds with

the experimental passage. Similar to Condition 1, the results show that there was an interaction between Duration and Frequency [ $F(4, 66) = 2.814, p = .032$ ] therefore, the main effect findings will not be reported. The effect size of this interaction effect was small (partial Eta-squared = .193). One finding from the interaction showed that for the altered stimuli in the 5% passage, listeners' ratings were significantly lower for 200 ms ( $M = 34.45$ ) compared to those for 300 ms ( $M = 72.52$ ) and 420 ms ( $M = 84.30$ ) durations. The mean ratings between 300 ms and 420 ms were not significantly different. Similar findings were also obtained for the 10% and 15% passages. The ratings were significantly lower for the 200 ms durations relative to the 300 ms and 420 ms durations but the ratings of 300 ms and 420 ms durations did not differ significantly.

### 3.3 Comparison of Listeners' Overall Speech Ratings Between Conditions 1 and 2

A comparison was made between the mean ratings of the overall speech of the speaker heard in Conditions 1 and 2. The mean differences in listeners' overall ratings in Condition 1 were subtracted from the overall ratings in Condition 2. Tables 3 and 4 show mean differences, standard deviations, and the SD/M ratios for the overall ratings in each condition. The average ratings are shown in Figure 3. Table 5 shows the paired sample  $t$  test analyses that were conducted between Condition 1 and 2 for each cell mean. To control for the inflated alpha level when multiple comparisons are made, the false discovery rate control method (Benjamini & Hochberg, 1995) was used, which resulted in an alpha level of .0278.

When the stimuli were manipulated in the 5% passage, mean difference ratings were lower in Condition 1 than Condition 2. However, there were no significant differences between the ratings for Conditions 1 and 2 at 200, 300, or 420 ms durations.

Stimuli altered in the 10% passage created a mean difference in listeners' ratings in Condition 1 that was descriptively higher than Condition 2 at the 200 ms duration ( $M_{\text{Condition 1-}}$

Condition 2) = .33). However, their ratings in Condition 1 were lower than Condition 2 for the 300 ms ( $M_{(\text{Condition 1-Condition 2})} = -16.58$ ) and 420 ms altered stimuli ( $M_{(\text{Condition 1-Condition 2})} = -17.25$ ). However, the results of paired samples  $t$  tests showed that there was no significant difference in mean ratings between Conditions 1 and 2 at 200 ms, but there were significant differences in mean ratings between Conditions 1 and 2 for the 300 ms ( $t(11) = -3.64, p = .004$ ), and the 420 ms altered stimuli ( $t(11) = -2.93, p = .014$ ).

In the 15% passage, mean differences in listeners' ratings were lower in Condition 1 than Condition 2 for 200 ms, 300 ms, and 420 ms altered stimuli. The results of paired samples  $t$  tests showed that there was no significant difference between Conditions 1 and 2 at 200 ms and 420 ms but there was a significant difference between the conditions for the 300 ms voiceless fricatives ( $t(11) = -3.19, p = .009$ ).

In summary, there were no statistically significant differences in listeners' ratings of the overall speech between Conditions 1 and 2 for any of the extended durations in the 5% passage. However, when the altered stimuli were presented in the 10% passages, listeners gave significantly higher ratings in Condition 2 than Condition 1 when the durations of altered sounds were 300 and 420 ms. In the 15% passage, listeners had significantly higher ratings between Conditions 1 and 2 when the altered stimuli were 300 ms in duration.

## 4. Discussion

### 4.1 The Effect of Phoneme Durations and Frequency on the Perception of Sound

#### Prolongations

One of the aims of this study was to determine if the identification and rating of words in passages as "stuttered" would be impacted by manipulations of the duration and frequency of occurrence of the initial voiceless fricatives of target words. The results provide empirical

evidence that both the duration of the altered sound and the frequency of occurrence of the altered phonemes within multiple paragraph length material had an impact on listeners' perception of a word containing a sound prolongation. Simply increasing the duration of the initial voiceless fricatives of specific words from 120 to 200 ms and presenting them 5% of the time in an experimental passage was sufficient to have a majority of the listeners (64%) identify and label target words as stuttered. Identification was even greater for the listeners who heard the 200 ms word-initial sounds in the 10% passage (95%). These results show that words can be detected as prolonged even when there is a small increase in the duration of a voiceless fricative and even more so when the frequency of their occurrence increases. This finding is similar to that of Hegde and Hartman (1976) who found that a majority of listeners judged single unit repetitions as stuttering when they occurred on 15% of the words in the speech sample.

The findings of the current study are also consistent with past studies that have tested changes in the perception of stuttering when manipulations of the duration and form of samples presented to listeners (Amir & Yairi, 2002; Curran & Hood, 1977; Jones et al., 2005; Kawai et al., 2005, 2007; Sander, 1963). Jones et al. (2005) found that the minimum sound duration perceived as abnormally long or stuttered was 235 ms for the duration when sounds began with /z/ and 279 ms when sounds began with /æ/. Given that the present study altered sounds at 300 ms, the results are more consistent with Lingwall and Bergstrand's (1979) study where listeners perceived voiced fricatives as stuttered when their durations were 294 ms or longer. The normal durations of voiced sounds are shorter than voiceless sounds; however, the thresholds that listeners perceived the sounds as prolonged were about the same between the current and previous studies. Therefore, the data from the present study extended our knowledge of listeners' perceptions of stuttering by showing that listeners perceive sounds within paragraph length

material as either abnormally prolonged or stuttered when the extended durations are approximately 300 ms, regardless of the type of phonemes.

The second major finding of this study relates to the listeners' ratings of the words in Condition 2. In this condition, listeners gave significantly lower ratings (i.e, toward the fluent end of the scale) for the 200 ms durations than the ratings for the 300 ms and 420 ms durations. This finding is similar to the discovery by Kawai et al. (2005, 2007) that ratings gradually changed from fluent to stuttered words when in the initial sounds of target words were systematically increased in duration. In their studies, listeners' mean ratings were 22.45 at 200 ms, 59.57 at 300 ms, and 74.03 at 420 ms. Using the same scale as in the Kawai et al. studies, the mean ratings for altered sounds in the current study were 51.15, 83.91, and 90.11 for 200, 300, and 420 ms, respectively. Comparing the ratings between studies, it is clear that listeners' perceptions of sound prolongations in the current study were much higher than in the previous studies by Kawai et al. On average, the ratings for the 200 ms sounds were perceived as half way between fluent and stuttered. However, once the sounds were 300 ms in duration, listener ratings showed that they perceived that the words resembled stuttering. The higher mean ratings in the present study could have been due to listeners being asked to rate words selected for them by the researchers or having an easier time rating words stuttered within the context of paragraph-length passages than in isolated sentences. Perhaps when a listener is presented with a connected speech context, there is more perceptual information to process and judge a speaker's degree of fluency.

One could argue that the findings of the present study were influenced by the labels (fluent vs. stuttered) and the rating scale (1-100) used in the study. It has been known for some time that the labels listeners are instructed to use when listening to various forms of nonfluent speech play a role in how those nonfluencies will be judged (Williams & Kent, 1958). In the

present study, it is unknown if the use of labels such as “fluent”, “nonfluent but not stuttered”, “nonfluent and somewhat like stuttering”, and “unambiguous stuttering” could have changed the ratings our listeners made during the experiment. Given that we had used the 1-100 rating scale in two previous studies (Kawai, Healey, & Carrell, 2005, 2007) to show that stuttering is perceived along a continuum, it was thought that the same scale would allow for a broad estimation of how words containing altered voiceless fricatives would be perceived. It is possible that our individual listeners might have interpreted the 1-100 scale differently such that a rating of “1” meant only fluent and a rating of “100” only represented stuttering, while others believed that giving a rating of 90 or more would indicate they perceived a stuttered event. The perceptual continuum used in this study did not allow for categories or labels beyond the two anchor terms. Thus, any value between the extremes of 1 and 100 is difficult to interpret and we acknowledge this limitation to our study.

#### **4.2 The Effect of Level of Attention to Duration of Altered Sounds on the Perception of Sound Prolongations**

This study is one of the first attempts to compare how listeners rate an overall or gestalt presence of stuttering when listening for prolonged sounds within a connected speech context (Condition 1) compared to their ratings of specific sounds within a passage that were selected for them by the researchers (Condition 2). In previous studies, listeners' perceptions of stuttering were tested without giving listeners specific instructions about which sounds or words to focus on (e.g., Schaeffer & Eichorn, 2001; Susca & Healey, 2001). Moreover, past research has used short, isolated phrases for listeners to evaluate during the experiment (e.g., Jones et al. 2005; Kawai et al., 2005). It seems reasonable to assume that listeners appear to rate stuttered words differently if they are instructed to rate specifically selected words as stuttered versus asking then



to select stuttered words on their own. Examining the theoretical implications of listeners' perceptions between two different perceptual levels of attention (i.e. focused vs. unfocused attention to stimuli) might shed light on this assumption.

Although published models or theories of stuttering do not address listener perceptions of stuttering, the current results can be explained within the context of Treisman's Feature-Integration theory (Treisman, 1986; Treisman & Gelade, 1980; Treisman & Schmidt, 1982) described earlier in the Introduction. The original feature integration theory applied to visual perceptions. However, many researchers have reported that feature integration theory is applicable for auditory perceptions of speech (e.g., Deutsch, 1986; Hall, Pastore, Acker, & Huang, 2000; Thompson, 1994; Thompson & Sinclair, 1993; Thompson, Hall, & Pressing, 2001) such as the types of listener perceptions that took place in the current study.

In the current study, Condition 1 measured the pre-attentive stage because participants were instructed to simply listen to the experimental passage and identify any stuttered words. In contrast, Condition 2 measured a listener's focused attention because they were directed to carefully listen to certain words that did or did not include altered stimuli. Thus, when the participants were asked to rate the speaker's overall fluency in Condition 1, they might have been in an unfocused stage because they were attending to content of the passage, whereas in Condition 2, listeners were asked to focus their attention and rate specific words within the passage. Comparing results across passages in Condition 2, listeners identified a higher frequency of the target words as stuttered in the passage with 15% altered stimuli compared to 5% altered stimuli. Additionally, participants who listened to the experimental passage at 10 and 15% frequencies of altered sounds gave higher ratings for Condition 2 than Condition 1 when the durations were 300 and 420 ms. However, when the passage was presented with 5% altered

stimuli, listeners' ratings were not significantly different between the pre-attentive and focused attention stages.

From a clinical perspective, what a listener focuses on could make a difference in whether the speech of an individual is considered stutter-like. For example, if parents are asked to listen to specific words in their child's speech vs. to tell whether any words within the connected speech sample sounded like prolongations, one might get different responses from the parents. A clinician might ask the parents to compare what they hear in the overall speech pattern versus their focused attention to specific disfluent words in their child's speech. Perhaps this information might be used to counsel the family about the child's stuttering.

#### **4.3 Future Research**

Although this is the first study to show that durations of altered sounds at different frequencies of occurrence in paragraph-length material influence listeners' perceptions of stuttering, additional studies are needed to assess the validity of the findings as well as the reliability of listeners' identification of stuttered words and ratings of stuttering on a scale of 1-100. In light of the current findings, future studies should use a larger number of participants in each listening condition and a broader range of listeners in terms of ages and different backgrounds. Additionally, researchers might want to consider manipulating different phonemes (e.g., vowels, stop consonants, voiced fricatives), using a larger number of sound durations, and possibly changing the methodology in terms of the ordering the conditions. Research along these lines is currently in progress.

Future studies in this area might also want to consider paying close attention to the location of strong and weak phonetically stressed syllables within an experimental passage. The level of syllable stress on words was not taken into consideration when creating the durations of

the voiceless fricatives although Umeda (1977) showed that a voiceless fricative within a stressed syllable is slightly shorter in duration than when the sound appears in an unstressed syllable. Our focus was on making all voiceless fricatives a standard length of 120 msec but it is possible that by doing so, small differences in listener judgments of stuttered words could have been affected by our failure to account for stressed and unstressed syllables. Moreover, efforts were made to vary where the target words appeared within the sentences such that some appeared at the beginning, middle, and end of the sentence. In retrospect, it is possible that changes in the position of the target words in the sentence could have also affected listener judgments of stuttered words. Finally, future studies might want to manipulate the speech characteristics of the speaker reading the experimental passage if slightly slower or fast rates of speech and small variations in the prosody of the speech sample could impact listeners' judgments of stuttering when listening to paragraph length speech material.

## 5. Conclusions

One aim of this study was to determine how well listeners identify words beginning with digitally altered voiceless fricatives as fluent or stuttered when produced at varying frequencies of occurrence within a passage. A second aim of this study was to determine if listeners would rate words as fluent or stuttered when they were and were not asked to focus on particular target words in the paragraph material. Target words within the passage were extended to 200, 300, and 420 ms in duration from their original durations of 120 ms. Each listener heard 5%, 10%, or 15% of altered stimuli at 200, 300, and 420 ms. The results showed that in Condition 1, a majority of listeners were able to identify stuttered words even when the 5% of the duration of the prolonged voiceless fricatives in the paragraph-length material were extended from 120 to 200 ms. This shows that even a small increase in the duration of sound at the beginning of a word can evoke

perceptions of stuttered speech. The results also showed that listeners perceived that words sounded like stuttering when their attention was directed to the evaluation of specific target words in the experimental paragraph. Thus, when listeners focus on certain words within a connected speech context, they tend to perceive stuttering even if the words in the sample occur 5% of the time and when their attention is drawn to particular words in the speech context. These results serve to advance our understanding of factors that influence listener perceptions of stuttering.

## References

- Amir, O. & Yairi, E. (2002). The effect of temporal manipulation on the perception of disfluencies as normal or stuttering. *Journal of Communication Disorders*, 35, 63-82.
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 298-300.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). Quantitative methods in psychology. *Psychological Bulletin*, 112, 155-159
- Curran, M. F. & Hood, S. B. (1977). Listener ratings of severity for specific disfluency types in children. *Journal of Fluency Disorders*, 2, 87-97.
- Deutsch, D. (1986). *Handbook of perception and human performance* (Vol. 2). New York: Wiley-Interscience.
- Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature-integration with spatially distributed items. *Perceptions and Psychophysics*, 62, 1243-1257.
- Hedge, M. N. & Hartman, D. E. (1979). Factors affecting judgments of fluency: II. Word repetitions. *Journal of Fluency Disorders*, 4, 13-22.
- Jones, K., Logan, K.J., & Shrivastav, R. (2005, November). *Duration, rate, and phoneme-type effects on listeners' judgments of prolongations*. Poster session presented at the meeting of American Speech-Language-Hearing Association, San Diego, CA.
- Kawai, N. (2007). The Pro Slider (Version 2.0). Lincoln, NE: University of Nebraska-Lincoln Speech Perception Laboratory.

- 1
- 2
- 3
- 4 Kawai, N., Healey, E. C., and Carrell, T. D. (2005, November). *Identification and discrimination*
- 5 *of phoneme prolongation*. Poster session presented at the meeting of American Speech-
- 6
- 7
- 8
- 9 Language-Hearing Association, San Diego, CA.
- 10
- 11 Kawai, N., Healey, E. C., and Carrell, T. D. (2007). Listeners' identification and discrimination
- 12
- 13
- 14 of digitally manipulated sounds as prolongations. *Journal of the Acoustical Society of*
- 15
- 16 *America*, 122, 1102-1110.
- 17
- 18
- 19 Klatt, D. H. (1973). Duration characteristics of pre-stressed word-initial consonant clusters in
- 20
- 21 English. *Quarterly Progress Report, Research Laboratory of Electronics MIT*, 108, 253-
- 22
- 23 260.
- 24
- 25
- 26 Lingwall, J. B. and Bergstrand, G. G. (1979, November). *Perceptual boundaries for judgments*
- 27
- 28 *of 'normal,' 'abnormal,' and 'stuttered' prolongations*. Poster session presented at the
- 29
- 30 meeting of American Speech-Language-Hearing Association, Atlanta, GA.
- 31
- 32
- 33
- 34 Murphy, C. (1989). Once upon a time. In *Challenger adult reading series* (Vol. 8, pp. 68-79).
- 35
- 36 Syracuse, NY: New Readers Press.
- 37
- 38 Sander, E. K. (1963). Frequency of syllable repetition and "stutterer" judgment. *Journal of*
- 39
- 40 *Speech and Hearing Disorders*, 28, 19-30.
- 41
- 42
- 43 Schaeffer, N. & Eichorn, N. (2001). The effects of differential vowel prolongations on
- 44
- 45
- 46 perceptions of speech naturalness. *Journal of Fluency Disorders*, 26, 335-348.
- 47
- 48 Sound Forge (Version 4.0c) [Computer software]. Madison, WI: Sonic Foundry
- 49
- 50
- 51 Srinivasan, N. & Carrell, T. D. (2004). Maketape (Version 2.2). Lincoln, NE: University of
- 52
- 53 Nebraska-Lincoln Speech Perception Laboratory.
- 54
- 55
- 56 Susca, M. & Healey, E. C. (2001). Perceptions of simulated stuttering and fluency. *Journal of*
- 57
- 58 *Speech, Language, and Hearing Research*, 44, 61-72.
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 1
- 2
- 3
- 4 Thompson, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration,
- 5
- 6 and pitch pattern with durational pattern. *Perception and Psychophysics* 56, 363-374.
- 7
- 8
- 9 Thompson, W. F., Hall, M. D., and Pressing, J. (2001). Illusory conjunctions of pitch and
- 10
- 11 duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human*
- 12
- 13 *Perception and Performance*, 27, 128-140.
- 14
- 15
- 16 Thompson, W. F., & Sinclair, D. (1993). Pitch pattern, durational pattern, and timbre: A study of
- 17
- 18 the perceptual integration of auditory qualities. *Psychomusicology* 12, 3-21.
- 19
- 20
- 21 Treisman, A. (1986). Properties, parts, and objects. In K. Boff, L. Kaufman, & J. Thomas (Eds.),
- 22
- 23 *Handbook of perception and human performance: Cognitive processes and performance*,
- 24
- 25 (Vol. 2, pp. 35-70). Chichester, UK: Wiley.
- 26
- 27
- 28 Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive*
- 29
- 30 *Psychology*, 12, 97-136.
- 31
- 32
- 33 Treisman, A., & Schmidt, N. (1982). Illusory conjunctions in the perception of objects. *Cognitive*
- 34
- 35 *Psychology*, 14, 107-141.
- 36
- 37
- 38 Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of*
- 39
- 40 *America*, 61, 846-858.
- 41
- 42
- 43 Walker, V. G. (1988). Durational characteristics of young adults during speaking and reading
- 44
- 45 tasks. *International Journal of Phoniatrics, Speech Therapy and Communication*
- 46
- 47 *Pathology*, 40, 12-20.
- 48
- 49
- 50 Williams, D., & Kent, L. (1958). Listener evaluations of speech interruptions. *Journal of*
- 51
- 52 *Speech and Hearing Research*, 1, 124-131.
- 53
- 54
- 55 Zebrowski, P. M. (1991). Duration of the speech disfluencies of beginning stutterers. *Journal of*
- 56
- 57 *Speech and Hearing Research*, 34, 483-491.
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

Zebrowski, P. M. (1994). Duration of sound prolongation and sound/syllable repetition in children who stutter: Preliminary observations. *Journal of Speech and Hearing Research*, 37, 254-263.



Table 1

*Means, Standard Deviations, and SD/M Ratios, for Participants' Percentages of Identifications of Duration Altered Speech in Condition 1*

Duration Frequency	200 ms	300 ms	420 ms
5%			
<i>M</i>	67.4	94.7	97.7
<i>SD</i>	25.8	6.1	5.7
<i>SD/M Ratio</i>	.38	.06	.06
10%			
<i>M</i>	84.5	99.3	99.6
<i>SD</i>	11.9	1.8	1.3
<i>SD/M Ratio</i>	.14	.02	.01
15%			
<i>M</i>	58.9	89.5	99.5
<i>SD</i>	27.6	15.9	1.2
<i>SD/M Ratio</i>	.47	.18	.01

Table 2

*Means, Standard Deviations, and SD/M Ratios of Participants' Ratings of Words at Each*

*Duration and Frequency for Condition 2*

Duration \ Frequency	Duration		
	200 ms	300 ms	420 ms
5%			
<i>M</i>	34.45	72.52	84.31
<i>SD</i>	8.02	6.72	12.41
<i>SD/M Ratio</i>	.23	.09	.15
10%			
<i>M</i>	56.84	87.57	91.38
<i>SD</i>	14.83	10.42	10.40
<i>SD/M Ratio</i>	.26	.12	.11
15%			
<i>M</i>	62.16	91.65	94.66
<i>SD</i>	11.87	5.85	5.97
<i>SD/M Ratio</i>	.19	.06	.06

Table 3

*Means, Standard Deviations, and SD/M Ratios for the Ratings of the Overall Speech in the "Fossils" Passage in Condition 1*

Duration Frequency	200 ms	300 ms	420 ms
5%			
<i>M</i>	21.83	29.17	39.42
<i>SD</i>	10.48	17.26	21.68
<i>SD/M</i> Ratio	.48	.59	.55
10%			
<i>M</i>	44.17	53.42	57.08
<i>SD</i>	19.43	20.12	24.33
<i>SD/M</i> Ratio	.44	.38	.43
15%			
<i>M</i>	42.83	53.25	64.75
<i>SD</i>	20.79	23.43	22.68
<i>SD/M</i> Ratio	.49	.44	.35

Table 4

*Means, Standard Deviations, and SD/M Ratios for the Ratings of the Speaker's Overall Speech in the "Fossils" Passage in Condition 2*

Duration Frequency	200 ms	300 ms	420 ms
5%			
<i>M</i>	29.17	35.17	40.00
<i>SD</i>	19.21	23.16	22.74
<i>SD/M</i> Ratio	.66	.66	.57
10%			
<i>M</i>	43.83	70.00	74.33
<i>SD</i>	23.92	16.93	20.23
<i>SD/M</i> Ratio	.55	.24	.27
15%			
<i>M</i>	48.83	75.75	80.42
<i>SD</i>	22.45	11.60	14.23
<i>SD/M</i> Ratio	.46	.15	.18

Table 5

*Results of Paired Samples  $t$  Tests between Condition 1 and 2 in Each Duration and Frequency Condition (Comparisons of Cell Means between Condition 1 and 2)*

Freq	Condition 1 -Condition 2	Paired Differences			$t$	$df$	$p$ (2-tailed)
		$M$	$SD$	$SEM$			
5%	200 ms	-7.33	20.37	5.8793	-1.25	11	.238
	300 ms	-6.00	29.82	8.6085	-.70	11	.500
	420 ms	-.58	28.15	8.1254	-.07	11	.944
10%	200 ms	.33	28.08	8.1057	.04	11	.968
	300 ms	-16.58	15.80	4.5601	-3.64	11	.004
	420 ms	-17.25	20.41	5.8906	-2.93	11	.014
15%	200 ms	-6.00	30.41	8.7784	-.68	11	.508
	300 ms	-22.50	24.40	7.0448	-3.19	11	.009
	420 ms	-15.67	24.26	7.0025	-2.24	11	.047

Figure Caption

*Figure 1.* Listeners' average percentage identified stuttered words in a passage who listened to either 5, 10, or 15% passages when the durations of altered sounds were 200, 300, and 420 ms. The bars represent observation means and the lines represent the 95% confidence intervals surrounding them.

Figure Caption

*Figure 2.* Listeners' mean ratings for altered sounds who listened to either 5, 10, or 15% passages when the durations of altered sounds were 200, 300, and 420 ms. The bars represent observation means and the lines represent the 95% confidence intervals surrounding them.

Figure Caption

*Figure 3.* Listeners' ratings of the speaker's overall speech in Condition 1 and Condition 2. The bars represent observation means and the lines represent the 95% confidence intervals surrounding them.



## Appendix A

### Fossils

Across centuries, many things have changed during the course of time since life began. For<sup>5, 10, 15%</sup> instance, the positions and shapes<sup>10, 15%</sup> of rivers have been changed dramatically. Animals and plants have changed so<sup>15%</sup> much, one would think<sup>5, 10, 15%</sup> none of us are from the same<sup>10, 15%</sup> planet. In order to understand fossils<sup>15%</sup>, we need to go back to a former<sup>5, 10, 15%</sup> geological era.

Our knowledge of the far distant past has been gained from a large amount of research on fossils. A physician<sup>5, 10, 15%</sup> by the name of George Sicola<sup>10, 15%</sup> was one of the first<sup>15%</sup> persons to use the word fossil<sup>10, 15%</sup>. He derived or formed<sup>5, 10, 15%</sup> the word fossil from the Latin word fossus<sup>5, 10, 15%</sup> that mean *dug*, meaning something<sup>15%</sup> dug out of the soil<sup>5, 10, 15%</sup>. The word fossil<sup>10, 15%</sup> came to be used for objects found<sup>15%</sup> in the earth or rocks that contained the remains of living things<sup>5, 10, 15%</sup>.

To be a fossil<sup>5, 10, 15%</sup>, a living thing<sup>10, 15%</sup> must be covered or buried extremely soon<sup>10, 15%</sup> after its death so<sup>15%</sup> oxygen can't reach it. Such<sup>5, 10, 15%</sup> burials may take place in sands<sup>15%</sup> and bogs. In dry regions, sandstorms bury and preserve animals. For<sup>10, 15%</sup> example, in regions of Central<sup>15%</sup> Asia, herds of animals ceased<sup>10, 15%</sup> to exist because of fierce<sup>15%</sup> desert storms. After sinking<sup>15%</sup> in water, some<sup>10, 15%</sup> animals are capable of becoming fossilized<sup>15%</sup> but these circumstances<sup>5, 10, 15%</sup> make it more difficult for fossils to form<sup>10, 15%</sup>.

Figure 1

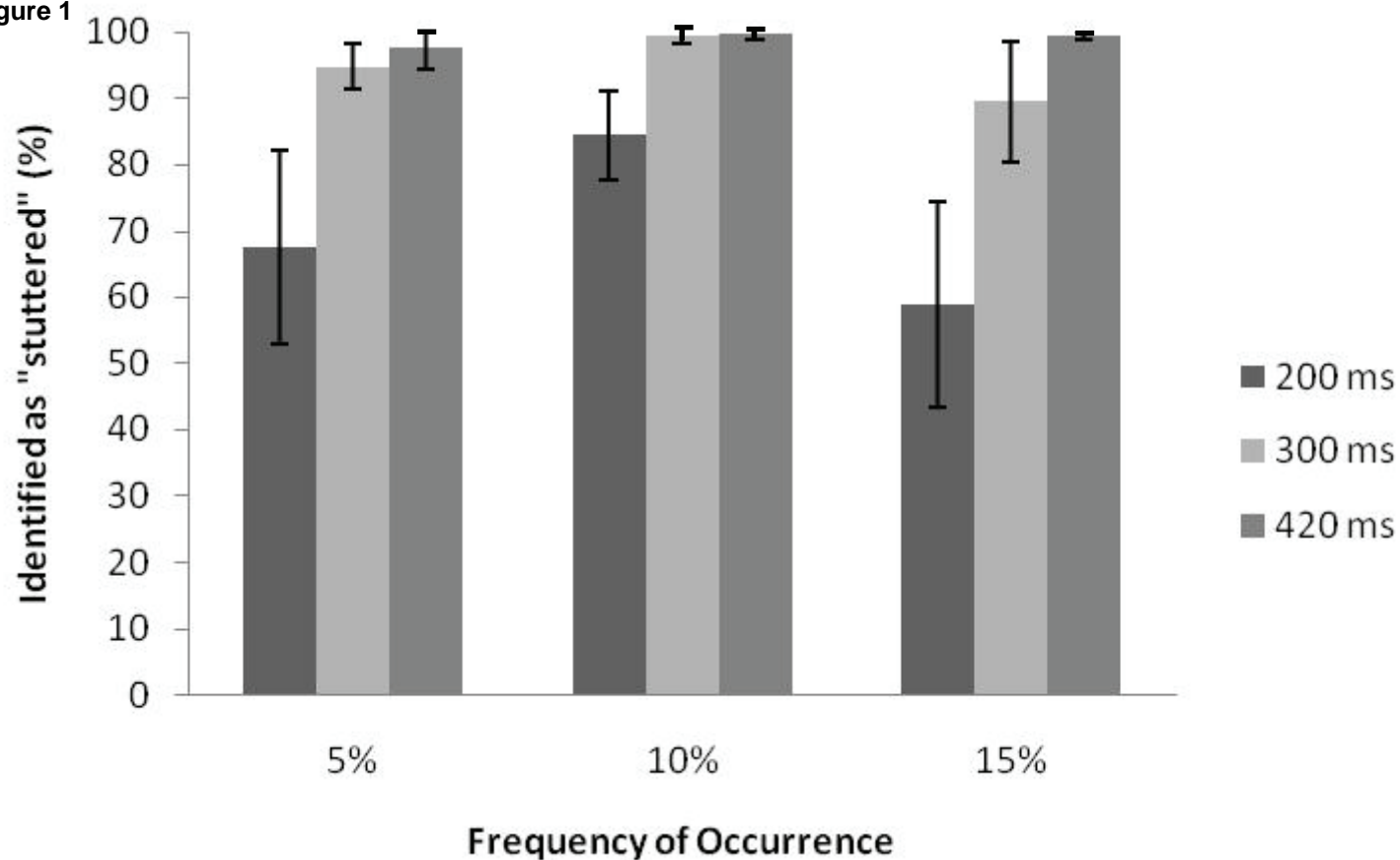
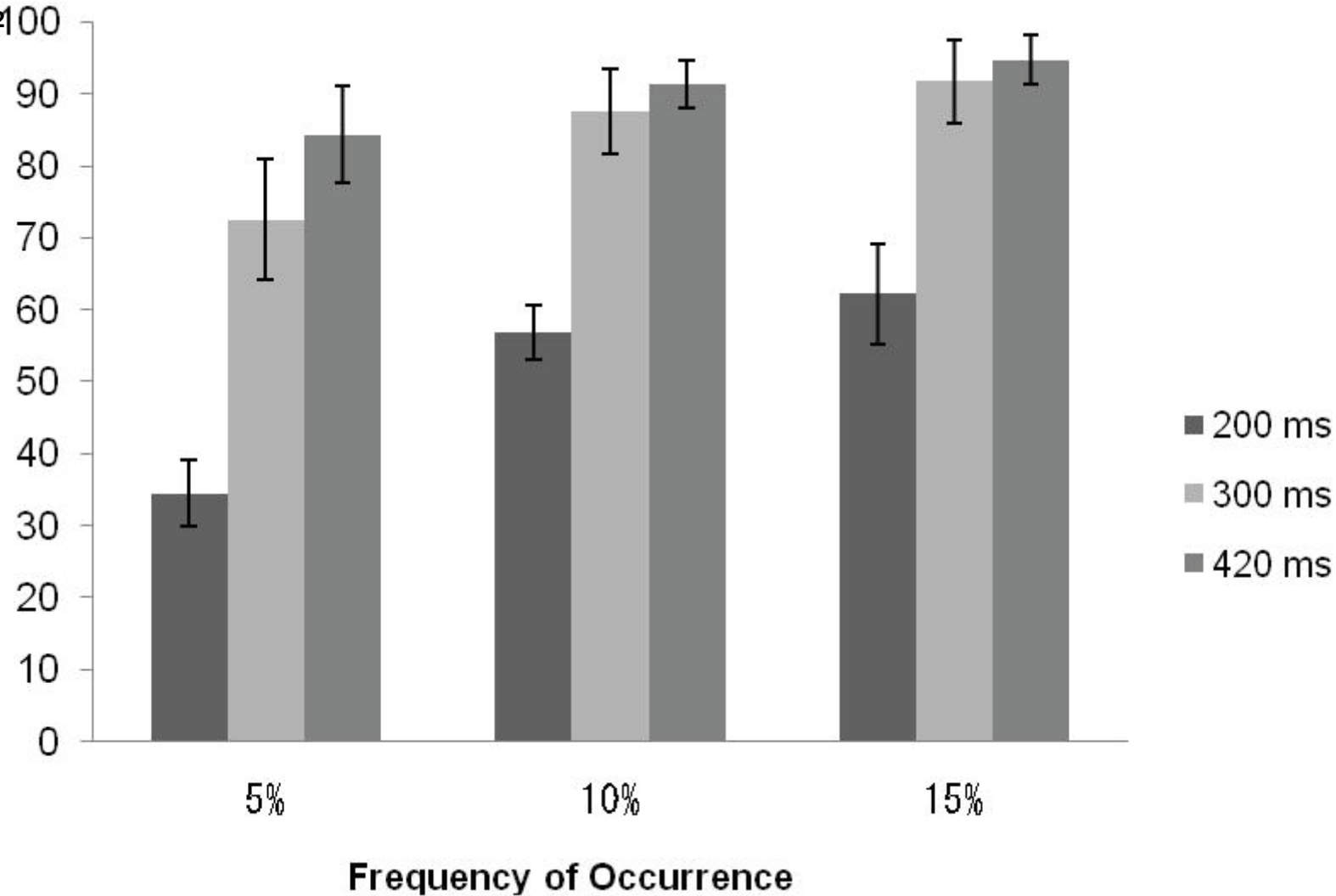
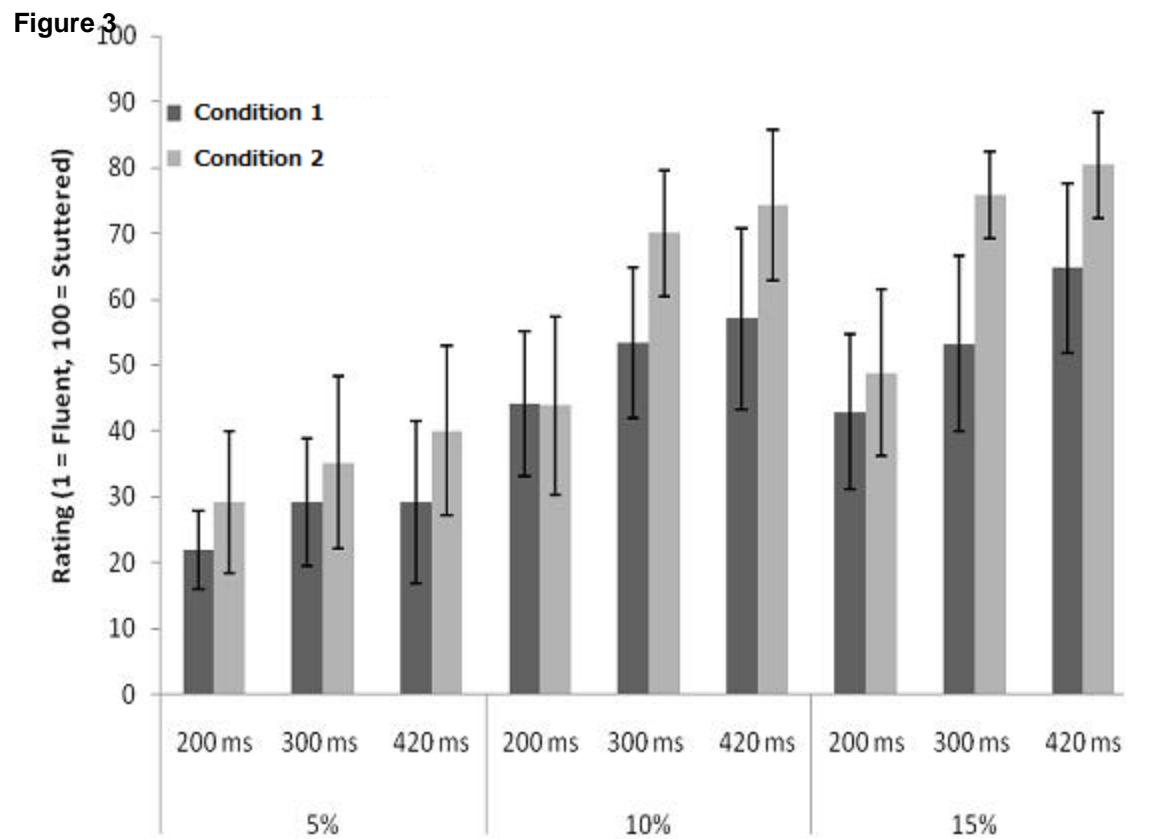


Figure 2

Ratings (1 = Fluent, 100 = Stuttered)





## \*Research Highlights

>We examined listeners' accuracy of the identification and rating of prolongations. >In Condition 1, listeners identified words considered stuttered. >In Condition 2, listeners rated specific words within the passage. >Both the duration & frequency of altered phonemes influenced perceiving stuttering. >Listeners rated higher in Condition 2 than 1 when words were stuttered or fluent.

## Appendix B. Continue Education Questions

1. Based on this article, the reason voiceless fricatives were chosen as the target experiment stimuli of the current study was because:
  - (a) it was easy to digitally extend the duration of the frication noise using a speech analysis computer program.
  - (b) the normal duration of voiceless fricatives are approximately the same.
  - (c) word-initial voiceless fricatives could be produced as sound prolongations by PWS.
  - (d) a and c only
  - (e) a, b and c
  
2. The reason only one frequency (5, 10, or 15%) of altered stimuli was presented to listeners was because:
  - (a) they tend to change their perceptions because of fatigue when listening to the same passage at different frequencies of occurrence.
  - (b) if they heard the 15% passage first, their perceptions would have been biased when they listened later to the 5% passage.
  - (c) there were limitations of the software (Maketape ver. 2.2) in presenting stimuli at different frequencies of occurrence.
  - (d) it would have been difficult statistically to separate the effects of the duration of the altered stimuli from the frequency at which they were presented.
  - (e) none of the above.
  
3. In Condition 1, listeners' percentage of identified stuttered words was:
  - (a) significantly smaller for the 200 ms than for 300 and 420 ms durations in the 5% passage
  - (b) significantly smaller for the 200 ms than for 300 and 420 ms durations in the 5 and 15% passages
  - (c) significantly smaller for the 200 ms than for 300 and 420 ms durations in all frequency-altered passages

- (d) significantly greater for the 200 ms than for 300 ms duration in all frequency-altered passages.
- (e) not significantly different among 200, 300, and 420 ms durations for any of the frequency-altered passages

4. In Condition 2, listeners' ratings of the target words:

- (a) only sounded like stuttering when the voiceless fricative was 420 ms in the 15% passage
- (b) were not sufficiently long enough at 200 ms to cause listeners to rate the word as stuttered
- (c) were rated as stuttered as long as they heard any duration of altered stimuli in the 15% passage
- (d) resembled a stuttered word when the altered duration reached 300 ms. for all three experimental passages
- (e) were similar across all durations and frequencies of altered stimuli

5. Listeners' ratings of target words as stuttered or fluent between Condition 1 and 2:

- (a) were not significantly different when the altered stimuli were presented in the 5% passage.
- (b) were significantly different for the 300 ms and 420 ms altered stimuli in the 10% passage
- (c) were significantly different for the 300 ms altered stimuli in the 15% passage
- (d) were significantly different at 200, 300, and 420 ms durations for all three experimental passages
- (e) a, b, and c above but not d.

Answers:

1. e
2. b
3. b
4. d
5. e



After reading this article, the learning will be able to:

1. describe the past literature on listener perceptions of stuttering
2. differentiate between listener's perceptions of sound prolongations that are altered in duration and frequency of occurrence.
3. describe how paragraph-length speech material compares to past research that has used isolated utterances