MEASURING CHANGES IN COGNITIVE LOAD IN LANGUAGE PROCESSING
THROUGH MEASURING POSTURAL SWAY

BY

JOHN JANG

DISSEbATION
Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:
Professor Chilin Shih, Chair and Director of Research
Professor Jacob Sosnoff
Assistant Professor Xun Yan
Assistant Professor Manuel Hernandez
Abstract

This thesis investigates the potential for using standing balance as a way to measure changes in cognitive load during language processing. Changes in standing balance are measured as subjects process vocabulary words, sentences, and paragraphs with varying degrees of relative difficulty. It is predicted that the more difficult the content, the more subjects will sway.

Although standing itself is a common enough task, the biological requirements of postural control during standing are fairly complex, requiring the involvement of an uninterrupted flow of afferent signals reaching the central nervous system from the muscle, tendon and joint proprioceptors, skin exteroceptors, and vestibular and visual inputs ([Fitzpatrick and McCloskey, 1994]). As such, there are numerous clinical applications in assessing standing balance, as decreasingly stable measured parameters from a standing posture can reflect the increased risk of falls in the elderly ([Maki et al., 1994]), individuals with Parkinson’s Disease ([Błaszczyk et al., 2007]), or patients with multiple sclerosis ([Prosperini and Pozzilli, 2013]). Standing balance is sensitive to psychological perturbations ([Kerr et al., 1985]), and may even be perturbed by language tasks ([Shumway-Cook et al., 1997]), indicating a link between standing and language processing.

Based on prior literature, this thesis tests the assumption that there is a link between language processing and postural control. If standing control processes are interrupted by the processing of language, then using balance to measure changes in cognitive load could be used to objectively assess linguistic difficulty.

Force plates are used to measure balance, which are laboratory-grade devices used in Biomechanics and Kinesiology to quantify various physical properties of human movement ([Winter, 2009]), ranging from running or jumping ([Chappell et al., 2005], [McLean et al., 2004]), to the more obscure and less conscious overall motions such as the subtle displacements involved in maintaining a standing upright posture ([Murray et al., 1975]).

A set of experiments was performed on the auditory processing of three general elements of language: Paragraphs, sentences, and vocabulary words. There are two methods of analysis used, one which tests the overall changes in postural sway, and one which observes changes in postural sway over time. The two parameters used, respectively, are the fiftieth percentile of the radial range of the center of pressure, and the
radial distance of the center of pressure over time.

The results from the overall changes in postural sway suggest that balance is sensitive to changes in cognitive load due to language processing. For the paragraphs, results suggest that very difficult passages elicit greater changes in postural sway compared to the easier passages. Likewise, in the first of two experiments on sentences, sentences which the literature reports requiring relatively more cognitive load to process exhibit relatively greater changes in postural sway. For the second sentence study, results suggest a correlation between postural sway and the acceptability rankings of four different sentence types, so long as syllable count is accounted for. For the experiment on vocabulary words, increases in postural sway are observed when the difficulty level is increased.

The results from the observations of postural sway over time also suggest that balance is sensitive to changes in cognitive load due to language processing. For the paragraphs, the trends of postural sway are different for each of the three difficulty levels tested. For the first of two experiments on sentences, it is shown that the sway is increased in the parts of the sentences where more cognitive load is required. Likewise for the second sentence study, a similar trend in postural sway is found for the regions of time where the sentences should be less acceptable. For vocabulary words, an observable trend is also present, and only a trend exists, implying that a threshold for utilizing postural sway to intuitively measure changes in cognitive load due to language processing is reached in vocabulary words.

Overall, the results suggest that postural sway is sensitive to changes in cognitive load for paragraphs, sentences, and possibly vocabulary words. An intuitive interpretation of the results can be found for the studies on paragraphs and sentences, where greater amounts of sway is indicative of greater difficulty encountered during language processing. This trend in the data is present, but less apparent, for the study on the vocabulary words, providing some insight into how small the tested elements of language can be in getting an intuitive interpretation of changes in postural sway as it relates to changes in cognitive load during language processing. These results suggest that postural sway can be used as a practical tool for measuring changes in cognitive load during language processing.
To Father and Mother.
Acknowledgments

I would like to express my sincere gratitude to my advisor Chilin Shih, who has guided me along the way during the process of making this thesis. I have learned countless valuable lessons pertaining to the realm of linguistics research through her support and assistance. Furthermore, Professor Shih and my committee members have provided invaluable feedback and suggestions, which I am also indebted and grateful for. I also wish to thank Greg Kochanski for his effort and patience in helping to revise the paper.

Additionally, my previous advisor in engineering allowed me to use her facilities again for this research, and carrying out the balance experiments has flowed very smoothly from a logistics standpoint. But I wish to express a special thanks to Professor Hsiao-Wecksler not only because of that, but because as her student, she gave me the spark to foster a passion in biomechanics, and the ideas in this thesis could not have developed without her guidance and instruction over the years.

Last but not least, thank you to my close UIUC friends, colleagues, and family, for their belief in me and ardent devotion.
# Table of Contents

List of Tables ........................................... ix

List of Figures ........................................... x

List of Abbreviations ................................... xiii

List of Symbols ......................................... xiv

Chapter 1  Introduction .................................... 1

1.1  Overview ........................................... 1

1.1.1  Postural sway and thinking ....................... 2

1.1.2  The dual-task paradigm .......................... 2

1.1.3  Sway and language .............................. 3

1.1.4  The present goal: Measuring balance during language processing .............. 3

1.2  Methodology ......................................... 4

1.2.1  Force plates and the center of pressure .......... 4

1.2.2  Calculating the fiftieth percentile of the radial range of the COP ............. 7

1.2.3  Generalized additive models ..................... 9

1.2.4  Experiments on paragraphs, sentences, and vocabulary words .................. 11

Chapter 2  Changes in Balance Corresponding to the Aural Processing of Paragraphs 13

2.1  Introduction ......................................... 13

2.1.1  Background: Passage difficulty and reading ease scores ......................... 13

2.2  Methodology ......................................... 15

2.2.1  Materials: Passage selection and the utilization of SCIgen, and the creation of the final token set ......................................................... 15

2.2.2  Procedure ........................................ 26

2.2.3  Subjects .......................................... 28

2.3  Data Analysis ....................................... 28

2.3.1  Results from the fiftieth percentile of the radial range of the COP and the radial distance data over time ............................................. 28

2.3.2  Discussion of the observed trends between postural sway and paragraph difficulty level ................................................................. 31

2.3.3  Summary of the paragraphs experiment ................................................. 32

Chapter 3  Changes in Balance Corresponding to the Aural Processing of Single Sentences 33

3.1  Introduction ......................................... 33

3.1.1  Background: Choosing tightly controlled sentence pairs ......................... 33

3.1.2  HNPS and SS/SSO Relative Clauses ............... 33

3.2  Methodology ......................................... 35

3.2.1  Materials: Creating a token set controlled for with respect to possible word effects .... 36

3.2.2  Procedure ........................................ 38

3.2.3  Subjects .......................................... 38
3.3 Data Analysis .................................................. 39
  3.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial
       distance data over time .................................. 39
3.4 Discussion .................................................... 39
  3.4.1 Discussion of the observed trends between postural sway and sentences .......... 39
3.5 Summary of the sentences experiment .............................................. 46

Chapter 4  Changes in Balance in a Replication Study on Acceptability Judgments for
Sentences .......................................................... 47
  4.1 Introduction .................................................. 47
    4.1.1 Background: Replication and verification of a study on the acceptability of sentences 47
  4.2 Methodology .................................................. 49
    4.2.1 Materials used from the study on acceptability ........................................ 49
    4.2.2 Procedure ................................................. 51
    4.2.3 Subjects .................................................... 51
  4.3 Data Analysis .................................................. 52
    4.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial
       distance data over time .................................. 52
    4.3.2 Discussion of the observed trends between postural sway and sentences .......... 52
    4.3.3 Factoring in syllable count ................................ 58
    4.3.4 A preliminary plot factoring in syllable count ...................................... 59
    4.3.5 A more formal analysis factoring in syllable count .................................. 61
    4.3.6 Summary of the verification study .............................................. 62

Chapter 5  Changes in Balance Corresponding to the Aural Processing of Vocabulary
Words ............................................................. 65
  5.1 Introduction: Choosing words under a set of controlled guidelines ................. 65
  5.2 Methodology .................................................. 66
    5.2.1 Materials: Creating the word token set by utilizing word frequency and Zipf’s Law . 66
    5.2.2 Subjects .................................................... 72
    5.2.3 Experimental procedure .................................... 72
  5.3 Data analysis .................................................. 73
    5.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial
       distance data over time .................................. 73
    5.3.2 Discussion of the relationship between postural sway and vocabulary words ....... 78

Chapter 6  Discussion and interpretation of the findings from the four studies on balance
and language ....................................................... 80
  6.1 Paragraphs study overall review ........................................ 80
    6.1.1 Paragraphs and the fiftieth percentile of the radial range of the COP .......... 80
    6.1.2 Paragraphs and the radial distance of postural sway using GAMs ............ 81
  6.2 HNPS and SS/SSO study overall review ...................................... 82
    6.2.1 HNPS and SS/SSO sentences and the fiftieth percentile of the radial range of the COP 82
    6.2.2 HNPS and SS/SSO sentences and the radial distance of postural sway using GAMs .... 83
  6.3 Verification study overall review ....................................... 84
    6.3.1 Verification study sentences and the fiftieth percentile of the radial range of the COP 84
    6.3.2 Verification study tokens and the radial distance of postural sway using GAMs .... 84
  6.4 Vocabulary study overall review ....................................... 86
    6.4.1 Vocabulary study and the fiftieth percentile of the radial range of the COP ..... 86
    6.4.2 Vocabulary study tokens and the radial distance of postural sway using GAMs .... 87
  6.5 Overall impact from the four studies ....................................... 89
    6.5.1 Testing paragraphs, sentences, and vocabulary words ............................. 89
    6.5.2 Vocabulary words as the limiting factor ........................................ 89
  6.6 Alternative methods of measuring changes in cognitive load during language processing .... 90
List of Tables

2.1 Numerical interpretations from the Flesch reading ease test. .......................... 14
2.2 Tokens for the easy level. ................................................................. 18
2.3 Tokens for the medium level. ............................................................. 21
2.4 Tokens for the difficult level. .............................................................. 23

3.1 HNPS token set, where the easier sentences are on the top set of rows, and the harder sentences on the bottom set of rows. ................................................................. 37
3.2 SS/SO token set, where the easier sentences are on the top set of rows (SS sentences), and the harder sentences on the bottom set of rows (SO sentences). ........................................ 37

4.1 Sentence tokens for the replication study. .................................................. 50
4.2 Tokens used from the replicated study has differing syllable counts across sentence types. 61

5.1 The final selection of the vocabulary words chosen, based on frequency levels. The table shows the vocabulary words chosen; associated frequency data is noted below each of the respective tokens, in words per million. Level is listed in the first column, and the respective upper and lower frequency tolerance ranges are listed below it. Nonsense words are listed in the last row, and have no associated frequency data. In column 4, the abbreviation "allit" next to the syllable pattern stands for vocabulary words containing alliteration. ........................................ 71
List of Figures

1.1 Subjects stand on a force plate, which allows the attainment of measures of balance for the individual. (To give a sense of dimensional scale, a regular U.S. dime has been placed on the upper right hand corner of the plate prior to taking the photo.) ............................................. 4
1.2 Demonstrative image to help conceptually understand the COP, where for gross movements, the COP will generally move in the overall direction for which a subject leans in. ......................... 5
1.3 Example of the COP plot for a given person’s standing trial. ................................................. 6
1.4 Plot of the COP trace for a given standing trial. The three circles represent the maximum, median, and minimum radial distances from the origin of the COP trace, labeled in red, orange, and green respectively. ................................................................. 8
1.5 Plot of the COP radial distance over time for a given standing trial. The three lines represent the maximum, median, and minimum radial distances from the origin of the COP trace, labeled in red, orange, and green respectively. ..................................................... 10

2.1 Standing in the tandem position. The dominant foot is placed directly in front of the non-dominant foot. ......................................................................................... 27
2.2 Plot showing that the log of the fiftieth percentile radial range of the COP is higher for the difficult level in relation to the medium and easy levels. ........................................... 29
2.3 Demonstrative figure representing the trend of the radial distance data in the paragraph tokens, where the x axis is the time axis, and the y axis is the radial distance. One token is chosen per difficulty level, which shows the general observed trend across difficulty level. Difficulty level is color coded such that blue is easy, yellow is medium, and red is difficult (for black and white printing, darker crisscross patterns indicate the harder difficulty level). The region of interest is the entire duration of the token. Over each difficulty level, a general trend is observed, where with regards to the overall slope of the radial distance over time, easy tokens are downwards-sloping, difficult tokens are flat-sloped, and medium tokens are somewhere in between. The more difficult the token is, the greater the slope of the radial distance becomes. For more detailed plots of all the tokens, see Figure 2.4. ......................... 30
2.4 A series of time plots that show the effect of paragraph difficulty on the slope of the radial distance of the COP. The plots are the result of fitted splines from the GAMs of the radial distance of the COP versus time, for each of the fifteen tokens. Difficulty is color coded from blue to red from the top row to the bottom row (for black and white printing, from light to dark crisscrosses). Easy passages are in the first row, medium passages in the second row, and difficult passages in the third row. ......................................................... 31

3.1 Both the HNPS and SS/SO sentence types have a relatively easy and difficult variant, from which the difference comes from only switching the word order. There are therefore a total of four sentence variants as shown in the plot, two of them being relatively easy, and two of them being relatively difficult compared to its respective counterpart. Each of the sentence type’s difficult variants are highlighted in a black crisscross pattern, while the relatively easier sentence variants are in clear white. ......................................................... 40
3.2 Demonstrative figure representing the trend of the radial distance data in the sentence tokens, where the x axis is the time axis, and the y axis is the radial distance. This token pair shows the general observed trend over the token's relative difficulty level. The only difference between a token pair is the word ordering. The easy token is on the left, and the hard token is on the right. The colored region indicates the region in time where the word order between the two tokens differ, such that the blue region is for the easier token, and the red color is for the hard region (for black and white printing, the harder region is shaded in a crisscross pattern). The dark black vertical line at the end of the colored region indicates where the tokens' respective difference in word order has finished. The region of interest has been circled, which is the region of time after the colored region. After the colored region, the more difficult variant (plot on the right) has a higher uptick in sway than the easier variant (plot on the left). For more detailed plots of all the tokens, see Figure 3.4.

3.3 Demonstrative figure representing the trend of the radial distance data in the sentence tokens, where the x axis is the time axis, and the y axis is the radial distance. This token pair shows the general observed trend over the token’s relative difficulty level. The only difference between a token pair is the word ordering. The easy token is on the left, and the hard token is on the right. The colored region indicates the region in time where the word order between the two tokens differ, such that the blue region is for the easier token, and the red color is for the hard region (for black and white printing, the harder region is shaded in a crisscross pattern). The dark black vertical line at the end of the colored region indicates where the tokens' respective difference in word order has finished. The region of interest has been circled, which is the region of time after the colored region. After the colored region, the more difficult variant (plot on the right) has a higher uptick in sway than the easier variant (plot on the left). For more detailed plots of all the tokens, see Figure 3.4.

3.4 Resulting fitted splines from the GAMs of radial distance versus time, for the HNPS tokens. In this figure, only the word orders are different between the token pairs in each row. These plots are titled as unshifted (plots in the left column) and shifted (plots in the right column), where shifted has incorrectly structured word order, and unshifted preserves the more natural word order. The region of time where the word orders differ from each other are highlighted in different colors, where the easier token (unshifted) is shaded in blue, and the more difficult token (shifted) shaded in red. The thick vertical black bar represents the end of the colored region, ie when the difference between the two token pairs' word order is respectively completed.

3.5 Resulting fitted splines from the GAMs of radial distance versus time, for the SS/SO tokens. In this figure, only the word orders are different between the token pairs in each row. The region of time where the word orders differ from each other are highlighted in different colors, where the easier token (SS) is shaded in blue, and the harder token (SO) shaded in red. The thick vertical black bar represents the end of the colored region, ie when the difference between the two token pairs’ word order is respectively completed.

4.1 The tokens from four general sentence types were taken from a study by [Juzek, 2016], with the resulting plot of sway versus each of the four sentence types. The less acceptable sentence types are crisscrossed in the plot, and the more acceptable sentence types are in clear white.

4.2 Demonstrative figure representing the general trend observed for the alt-if tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the general trend. It is observed that the radial distance generally follows a downwards sloping trend over the entire token. For more detailed plots of all the tokens, see Figure 4.6.

4.3 Demonstrative figure representing the general trend observed for the resumptive tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. The region in time where the sentence can no longer become salvageable in terms of its grammaticality is highlighted in red. (For black and white printing, the red region is also shaded in a crisscross pattern.) When the subject listens to the sentence in this less acceptable region (circled above), there is a relative uptick in sway. For more detailed plots of all the tokens, see Figure 4.6.
4.4 Demonstrative figure representing the general trend observed for the good tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. It is observed that the radial distance generally follow a downwards sloping trend over the entire token. For more detailed plots of all the tokens, see Figure 4.6.

4.5 Demonstrative figure representing the general trend observed for the bad tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. It is observed that the radial distance fluctuates more relative to the trends in the other sentence types. For more detailed plots of all the tokens, see Figure 4.6.

4.6 Resulting fitted splines from the GAMs of radial distance versus time, for the tokens in the verification experiment. Parts where the sentence is expected to be easy are highlighted in blue, and hard highlighted in red. For black and white printing, the red region is also highlighted in dark crisscrosses.

4.7 Plot of the normalized fiftieth percentile of the radial range of the COP for the sentence types in the verification study, this time normalizing by the syllable count. The tokens from four general sentence types were taken from a study by [Juzek, 2016], with the resulting plot of sway versus each of the four sentence types. Postural sway data is normalized by subject and divided by the average syllable count of each of the tokens per sentence type. The less acceptable sentence types are crisscrossed in the plot, and the more acceptable sentence types are in clear white.

4.8 When syllable count is factored into the regression equation, the transformed sway data closely matches the numbers from the acceptability judgments with a slope very close to 1 at 0.946. The R squared value is also at .977, which further suggests that postural sway may have a relationship with syllable count.

5.1 Plotting the median frequencies by difficulty reveals a rapidly decreasing trend. However, creating a log-log plot of the figure on the left reveals a linearly decreasing trend. Frequency values for word level difficulty are computed by utilizing the best fit line equation from the figure on the right.

5.2 Plot showing the effect of word difficulty on postural sway. More difficult levels are shaded in a darker crisscross pattern, such that the more difficult the level, the darker the crisscross pattern. Words were chosen from various difficulty levels based off of the assumption that more difficult words occur less frequently. Words on level 1 are fairly easy (ie commonly occurring), and words on level 21 are more difficult (ie less frequently occurring).

5.3 Plot showing the effect of syllable structure on postural sway. More complex syllable structures are shaded in a darker crisscross pattern. "C" stands for consonant, and "V" stands for vowel, such that a word like "toe" for example has a structure of CV, "cat", has a structure of CVC, etc. Note that CVCVC (allit) is the alliterated version of the CVCVC syllable structure type.

5.4 Demonstrative figure representing the general trend observed for the vocabulary words. The x axis is time, and the y axis is the radial distance. One plot is chosen for each of the seven difficulty levels to illustrate the trend. The plots are arranged from top to bottom in order of increasing difficulty level. In looking at the seven plots, it is observed that as the difficulty level increases, the overall slope of the radial distance becomes more negative. For more detailed plots of all the tokens, see Figure 5.5.

5.5 Resulting spline fits from the GAMs for the radial distance of the postural sway over time. Each subplot represents a particular token. The syllable type is displayed in the columns (first column is CVCV, second column is CVCVC, third column is CVCVC (alliterated), fourth column is VCCVC, and fifth column is CCVCVC). The levels are displayed in the rows (first row is level 1, second row is level 3... last row is level 21). The levels are also shaded from blue to red, or in black and white printing from lighter to darker crisscross patterns.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COP</td>
<td>Center of Pressure.</td>
</tr>
<tr>
<td>SS</td>
<td>Subject-subject.</td>
</tr>
<tr>
<td>SO</td>
<td>Subject-object.</td>
</tr>
<tr>
<td>NP</td>
<td>Noun phrase.</td>
</tr>
<tr>
<td>HNPS</td>
<td>Heavy noun phrase shift.</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized additive model.</td>
</tr>
</tbody>
</table>
List of Symbols

\( \alpha \) Coefficient for ordinary least squares linear regression.
\( \beta \) Coefficient for ordinary least squares linear regression.
\( \epsilon \) Coefficient for ordinary least squares linear regression.
Chapter 1

Introduction

1.1 Overview

The goal of this thesis is to demonstrate that there is an underlying relationship between standing balance and language processing. The present work investigates changes in balance to find a simple and efficient way to measure changes in cognitive load without the need for introspection. If balance and language processing draw from the same pool of cognitive resources, then changes in postural sway would allow for an objective quantification of changes in cognitive load.

Therefore, the following studies in the main body of this thesis are designed with one overall purpose, where although every experiment focuses on a different language task, they all converge in pointing out the specific ways for which postural sway was actually affected per task. The findings from each of these studies provide evidence of the feasibility of utilizing postural sway to objectively observe changes in cognitive load during language processing.

This first chapter provides a brief overview in laying out the overall framework of the thesis and the motivation for studying postural sway as it relates to language processing. Also discussed is the overall background, experimental framework, and the reasons for performing the language experiments. The parameters used for the analysis of the postural sway data will also be discussed. The research studies themselves will consist of chapters split up as accordingly, where chapters 2 through 5 will be logically split up into the studies’ respective procedures, results, and conclusions. A comprehensive discussion of the results is presented in chapter 6, and lastly, chapter 7 will make concluding remarks and discuss potential applications.

Having explained the structural framework of the thesis, the following section will detail the theoretical groundwork, including the rationale for which the studies of the present thesis will entail.
1.1.1 Postural sway and thinking

In the seemingly simple act of maintaining an upright posture in standing, postural sway is never perfectly stationary ([Winter, 2009]). As common in daily life as it may be, the biological requirements of postural control during standing are fairly complex, requiring the involvement of an uninterrupted flow of afferent signals reaching the central nervous system from the muscle, tendon and joint proprioceptors, skin exteroceptors, and vestibular and visual inputs ([Fitzpatrick and McCloskey, 1994]). Standing requires the active, continual control of muscles all over an individual’s body, and it is a state of unstable equilibrium ([Balasubramaniam and Wing, 2002]). With regards to postural control, it seems that standing is not a simple task, as it is nearly impossible for humans to stand perfectly still, even if it is the only task required of us.

Furthermore, literature suggests that people move when they think, whether they are explicitly aware of it or not. This is demonstrated by several studies that have looked at postural sway while subjects perform various cognitive tasks ([Johansson et al., 1994], [Schmahmann et al., 2000]), which show that there is a shared link between cognition and balance; see also [Woollacott and Shumway-Cook, 2002] which demonstrates that memory tasks impact postural sway. Living in a gravitational environment, human posture and standing are continuous tasks that must be accomplished ([Winter, 2009]), and it is found, for example, that postural sway was affected as a consequence of performing the cognitive task such as counting ([Maylor and Wing, 1996]) or spatial processing ([Kerr et al., 1985]).

1.1.2 The dual-task paradigm

In the aforementioned studies that investigate balance during the performance of cognitive tasks, an experimental design known as the dual task paradigm is used. In this experimental design, subjects are instructed to perform two tasks simultaneously. If the two tasks draw from the same pool of cognitive resources, one or both tasks may be affected. For instance, tasks such as rubbing one’s belly and patting one’s head might interfere with each other, as would texting while driving. Dual task paradigms have been applied in many different ways and across various fields. Some more relevant examples involve verbal fluency with tone discrimination ([Mayer et al., 2005]), judgment with recall ([Hertzog et al., 2002]), or comprehension task with lexical decision task ([Burkhardt, 2005]). In the aforementioned studies involving counting ([Maylor and Wing, 1996]) and spatial processing ([Kerr et al., 1985]), the two general tasks involved were the cognitive task, and the act of maintaining a standing balance.

By utilizing the dual task paradigm, it has been shown that even language tasks can perturb balance.
1.1.3 Sway and language

Among the studies reporting a link between cognition and balance, most relevant for present purposes, one particular study ([Shumway-Cook et al., 1997]) focuses on language tasks, and reports significant degradation in postural sway when performing the two tasks of standing and sentence completion. In that study, subjects are instructed to stand on a force plate, and have thirty seconds to verbally complete partially-constrained four-word sentences. For example, if a computer monitor displays the text "*____ b____ *____ f____", subjects have to utter a four-word sentence based on the letter restrictions - in this case for example, the boy ran fast is one such sentence.

A significant degradation in balance is found for the subjects that perform this language task. With the design of the dual task paradigm, under the assumption that the language task and standing task both tapped into the same pool of cognitive resources, the researchers had demonstrated that postural sway was affected. These results appear promising in perspective of the goals of the present study.

1.1.4 The present goal: Measuring balance during language processing

The apparent link between language and postural sway has promising value in the realm of language research. If experiments are designed in accordance with the dual task paradigm, then postural sway can help assess various facets of language processing as it relates to cognitive load. Postural sway may give both an objective and elegant way to measure changes in cognitive load with regards to language processing: If the language task is more demanding, then people will sway more.

Overall, the potential benefit in the present research is relatively straightforward. For language studies, if postural sway is used as a way to measure changes in cognitive load, then the benefit gained would be that utilizing postural sway avoids introspection. Introspection, ie the self-examination of one’s own mental processes, would ideally be avoided in language research, because it relies on the subjective conversion of a thought or impression into a numerical value. Data collection using introspection in language research is efficient and economical. However, introspection has the disadvantage of being subjective, and subjectivity associated with introspection is an obstacle in language research that has to be avoided with rigorous research design. By understanding the relationship between balance and language processing, one may be able to objectively measure changes in cognitive load while still retaining the benefits of being both efficient and economical.

In the present thesis, the hope is to utilize standing balance in order to objectively measure the degree to which humans process various aspects of language. Any relative difficulties in language processing will be assessed in utilizing the dual task paradigm, therefore investigating the relationship between postural sway
and changes in cognitive load.

1.2 Methodology

1.2.1 Force plates and the center of pressure

The device that is most often used to accurately measure standing balance is known as a force plate (Figure 1.1). Subjects stand on the device, participate in a given experiment, and through analysis of the outputted data channels, the center of pressure (COP) can be calculated. The COP tells us where the subject is standing at in a particular point in time, and analyzing the COP over the duration of a trial can provide useful information about the subject’s balance.
Figure 1.2: Demonstrative image to help conceptually understand the COP, where for gross movements, the COP will generally move in the overall direction for which a subject leans in.

To give a rough but quick intuitive understanding, the COP can be thought of as approximating the shadow of the center of gravity when standing. For example, if a subject leans in a given direction, the COP will tend to move in that particular direction as well (Figure 1.2). More precisely, however, the COP refers to the location of the normal ground reaction force of the subject, for whichever action he or she is performing. For instance, in drawing a free-body-diagram of the subject when h/she is standing, the COP is the x-y coordinate location of the net resultant ground reaction force, which moves around over time and has subtle variations in movement as the subject stands. An example of the COP trace is shown in Figure 1.3. (Because of the seemingly random visual nature of the COP, the COP trace in the figure is often referred to as a ‘spaghetti’ plot.) Force plates are designed to capture subtle movements of the human body at sufficiently high sampling rates, ranging from 40Hz to 1000Hz or more. Before parameter calculation, COP data is typically smoothed with a low-pass filter, and in the present study, the center of pressure data is filtered with a 4th-order 10-Hz low-pass zero-lag Butterworth filter. Postural sway data is captured at 100Hz. From this COP trajectory data, a large variety of parameters can be calculated, but for the present study, a focus will be placed on two parameters to get a feel for how much the subject swayed, which is detailed in the following section.
Figure 1.3: Example of the COP plot for a given person’s standing trial.
1.2.2 Calculating the fiftieth percentile of the radial range of the COP

A multitude of parameters can be calculated from the COP data, such as COP velocity, standard deviation, or the range of the COP excursion. Of the parameters available, a focus will be placed on the radial range of the COP excursion, because in a previous pilot study regarding sentence processing and balance, the range of COP displacement in particular showed promising differences between sentences containing different types of grammatical errors ([Jang, 2010]). More specifically, a smaller value of the radial range of the COP will be used, the reasoning of which will be explained in the paragraphs below.

The radial range of the COP excursion is the difference in radii between two circles encompassing the COP trace. This would be the difference in value from the radius of the smallest circle containing the COP trace, subtracted from the radius of the largest circle containing none of the COP trace. In an idealized example, if a subject swayed with an outer radius of 20.0mm and an inner radius of 5.0 mm, the range of the COP would be 20.0-5.0=15.0mm.

In addition, however, for the analysis of the data in the following studies, a smaller percentage of the radial range of the COP will be used, which will be referred to in the present thesis as the fiftieth percentile of the radial range of the COP. This parameter is chosen in an attempt to reduce noise that may be present in the resulting data set from the experiments done in the present thesis. In other words, instead of using the outer circle, a smaller circle containing fiftieth percentile of the postural sway data is used instead. Put in more mathematical terms, the median value of the radial distance is taken, and that radius is considered to be the value for the outer circle. Another set of plots are displayed to help in understanding the physical calculation of fiftieth percentile radial sway parameter.

An additional and visual explanation of the fiftieth percentile of the radial sway is shown in Figure 1.4. This figure is used for visual explanation, and has a given subject’s COP while standing displayed for a period of time. Looking at the plot, one can see that the COP as shown in light gray roams rather freely, again demonstrating of its nickname as something of a spaghetti plot. If one were to quantify this subject’s overall balance using the COP in the spaghetti plot, there are several possible approaches. For instance, one could look at the maximum radius value as shown by the circle tagged with the red label and make corresponding interpretations. However, most of the person’s COP doesn’t lie along the maximum radius, and reporting the maximum value may skew one’s perspective of the subject’s overall balance. An average value could alternatively be calculated, but in that case a very high maximum value could also skew the average value and therefore the interpretation of the subject’s overall balance as well. Therefore, if one looks at the median value of the subject’s postural sway, as shown by the circle tagged with the orange label, one may be able to get a more stable and reliable understanding of the subject’s balance. Additionally, it is
noted that there is a region in the center of the spaghetti plot where the COP trace does not fall on. In other words, there is a region in the center of the plot, denoted by the ‘minimum’ tag highlighted in green, where the subject does not sway in. Thus, one can take the radius of the circle touching the median of the subject’s COP (tagged in orange), and the radius of the circle touching the minimum of the subject’s COP (tagged in green), and subtract the two radius values. The resulting value is shown in the orange region of the figure and can give a representative value of the subject’s balance, and this parameter will be referred to as the fiftieth percentile of the radial range of the COP.

To further help understand the meaning of this parameter, another visual aid of the calculation of the fiftieth percentile of the radial range of the COP is shown in Figure 1.5. The calculation is identical, only displayed using an alternative visual, where the radial distance of the subject’s COP is plotted against time. The figures represent the same person’s standing data, and the maximum, minimum, and median values
shown in Figure 1.5 match up with the radius data shown in Figure 1.4. It turns out in this trial, that the subject swayed his most near the beginning, as denoted by the red x, which is the same value as the circle’s radius labeled as ‘maximum’ in red from the previous figure. The median value was hit shortly beforehand and is denoted by the orange x, corresponding to the circle’s radius labeled as ‘median’ in orange from the previous figure. In this figure of radial distance plotted against time, the region covered below the horizontal line of the median value of the radial distance encompasses a good amount of the person’s sway, while at the same time avoiding high maximal values which may affect or bias an objective assessment of the individual’s overall balance. Additionally, there is a region below the minimum, this value denoted by a green x, where the subject does not sway in, and this corresponds to the circle’s radius labeled as ‘minimum’ in green from the previous figure. Therefore, if one subtracts the minimum value of radial distance from the median value of the radial distance, one ends up with the same parameter, ie the fiftieth percentile of the radial range of the COP.

1.2.3 Generalized additive models

In addition to the fiftieth percentile of the radial range of the COP (which are representative of a trial’s entire duration), time-related trends of the COP will also be looked at (ie, the COP radial distance black-line data shown in Figure 1.5). The reasoning for this time-related analysis is because in the experiments that are to be done in the present thesis, differences may also lie in the sway data during a specific word or phrase, the timing of which might not be captured by just looking at radial range data. In other words, while radial range gives a useful value for interpretation of standing balance, it is less informative about trends at specific points in time. To accommodate for time-related trends, one might wish to take sets of trials and average it over time, use a best-fit polynomial curve, etc., for which corresponding interpretations could be made.

For this thesis, generalized additive models (GAMs) are used to look for any telling trends of the COP radial distance data over time. GAMs attempt to explain a set of data using fitted spline functions. In other words, it takes inputted data and fits it with penalized cubic splines, and also returns a percentage value of the deviance that the model explains. Whereas mixed models attempt to explain a given dataset with straight line fits, GAMs use sets of connected and smooth polynomial fits instead.

GAMs have been utilized in a diverse number of research fields. For instance, GAMs are commonly used when investigating biological systems, such as species abundance due to environmental characteristics ([Forney et al., 2012]) and air quality effects on mortality ([Zhang et al., 2014]). GAMs have also have been used for predictions of bankruptcy ([Berg, 2007]), and lung cancer rate detection ([Clements et al., 2005]). Overall, GAMs are utilized to investigate nonlinear trends without prejudging what the exact shape of the
Figure 1.5: Plot of the COP radial distance over time for a given standing trial. The three lines represent the maximum, median, and minimum radial distances from the origin of the COP trace, labeled in red, orange, and green respectively.
outcome may be ([Crawley, 2012]), where the fitted response may more realistically represent the true response shape of the data ([Brigham, 2003]). GAMs offer rich exploratory flexibility ([Marx and Eilers, 1998]) and are used in exploratory analysis where the fitted splines allow for visualization of the relationship between the dependent and independent variables ([Xiang, 2001]). GAMs explains a specific percentage of the deviance of the experimental data, and there is no strict or hard-placed cutoff for the percentage of deviance explained, where for example studies having 15 percent of the deviance explained are reported in the literature ([Víkingsson et al., 2016]), and studies having 70 percent are also reported ([Zanobetti et al., 2000]).

To sum, for the studies in the present work, rather than look at each individual trial and make estimations about trends in the data, GAMs will fit the data instead, and as a result, the interpretations made from the time-related trends in the data will be more objective and practical. The primary reason for utilizing GAMs in the present study is to rely on one of the strengths of GAMs, which is to provide graphical results that speak for themselves. The graphical results present themselves as what they are such that visual observations to better understand the data can be reached.

1.2.4 Experiments on paragraphs, sentences, and vocabulary words

This thesis will attempt to demonstrate that language in general can invoke differing amounts of changes in postural sway when standing, depending on its relative difficulty level. In addition to full bodies of texts, there is an incentive to address the underlying and more fundamental building blocks of paragraphs, namely, sentences and vocabulary words. Therefore in this thesis, a focus is placed on postural sway as it relates to a) paragraphs, b) sentences, and c) vocabulary words.

All four experiments are performed in order to get a general feel for the extent of which postural sway is affected by language tasks. Testing just paragraphs would leave an uncertainty as to what the possible effect of individual sentences have on paragraphs during postural sway, and likewise the effect of words on sentences. The research will be performed on all three in order to at least get a sense of the extent to which postural sway may be affected by language.

The chosen order of paragraphs, sentences, and vocabulary words are done so as to get an intuitive grasp of the overall limit for which changes in postural sway may be affected by language tasks. Considering that paragraphs contain both sentences and words, and that sentences contain words, an opportunity presents itself to investigate the extent for which language tasks may affect postural sway. In the present thesis, the experiments are presented in a general order of decreasing complexity. Paragraphs contain more information than sentences, and sentences contain more information than words, and if a limit is reached in the experiments where no significant effects are found, a practical lesson about how far postural can be pushed could
be attained.

A focus placed exclusively on paragraphs is in the next chapter, followed later by two experiments on sentences, and then vocabulary words, as this setup will help maintain a clear path for the direction of the overall thesis. This next chapter of this thesis deals with postural sway as it relates to paragraphs, using basic readability equations for chosen text passages as a way to estimate paragraph difficulty. In this next chapter it is generally hoped that the more difficult the passage, the more people will think, which would lead to a general increase in the subject’s postural sway.
Chapter 2

Changes in Balance Corresponding to the Aural Processing of Paragraphs

2.1 Introduction

In this first study, language and standing balance is investigated, focusing on understanding subjects’ postural sway patterns while they process full paragraphs. If there is a link between changes postural sway and changes in cognitive load during language processing, one might intuitively expect that more difficult passages would elicit greater amounts of postural sway. Therefore, rather than use an arbitrary assortment of passages, a controlled set of passages are created that are split up by difficulty level, and the subjects’ postural sway response to the passages of varying difficulty levels is recorded.

The designed passages for the present experiment are split up into three levels of "easy", "medium", and "difficult". These passages are generally based off of sixth grade writing, high school writing, and journal-article-level writing, and are also controlled for by ensuring that the readability stays within a controlled numerical range using a basic equation for readability ease. As a result of careful consideration and control for establishing the three difficulty levels, data is collected and the resulting differences in postural sway is observed.

One of the measures for controlling the token difficulty is the Flesch reading ease formula, which will be discussed in the following section.

2.1.1 Background: Passage difficulty and reading ease scores

As mentioned in the previous section, the designed paragraph tokens fall into three separate difficulty levels, "easy", "medium", and "difficult". Postural sway from listening to these paragraphs is observed accordingly. In designing a passage of text for its intended audience, one factor worth considering is whether the overall readability of the passage is set at an appropriate level.

Various methods exist in the literature for which to assess appropriate levels of passage difficulty, and one way to think about different levels of passage difficulty is to relate it to educational grade level. This is a researched topic in the realm of childhood education ([Fisher et al., 2012]), and be-
cause instructors will utilize texts of differing complexity for their students to learn ([Fisher et al., 2012], [Hiebert and Mesmer, 2013]), there exists various methods to evaluate the text complexity of passages ([Mailloux et al., 1995], [Walsh and Volsko, 2008], [Ley and Florio, 1996]).

To this end, there are ways to numerically calculate a readability score, and the Flesch reading ease test ([Kincaid et al., 1975]) is one such method, which can roughly determine the readability of a given body of text. The Flesch reading ease test is an equation consisting of the following,

\[
206.835 - 1.015 \left( \frac{\text{totalwords}}{\text{totalsentences}} \right) - 84.6 \left( \frac{\text{totalsyllables}}{\text{totalwords}} \right)
\] (2.1)

where the words, syllables, and sentences are counted and placed into the above equation to give an estimate for how readable the passage may be (Table 2.1).

Table 2.1: Numerical interpretations from the Flesch reading ease test.

<table>
<thead>
<tr>
<th>score</th>
<th>school level</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.0 to 100.0</td>
<td>5th grade</td>
<td>Very easy to read. Easily understood by an average 11-year-old student.</td>
</tr>
<tr>
<td>80.0 to 90.0</td>
<td>6th grade</td>
<td>Easy to read. Conversational English for consumers.</td>
</tr>
<tr>
<td>70.0 to 80.0</td>
<td>7th grade</td>
<td>Fairly easy to read.</td>
</tr>
<tr>
<td>60.0 to 70.0</td>
<td>8th &amp; 9th grade</td>
<td>Plain English. Easily understood by 13- to 15-year-old students.</td>
</tr>
<tr>
<td>50.0 to 60.0</td>
<td>10th to 12th grade</td>
<td>Fairly difficult to read.</td>
</tr>
<tr>
<td>30.0 to 50.0</td>
<td>college</td>
<td>Difficult to read.</td>
</tr>
<tr>
<td>0.0 to 30.0</td>
<td>college graduate</td>
<td>Very difficult to read. Best understood by university graduates.</td>
</tr>
</tbody>
</table>

Readability equations such as the Flesch reading ease offer a way to look at readability of text in a generalized sense. If used properly, computing text readability can be utilized in real-life scenarios, where, for instance, readability is calculated for high school students needing to understand a well-constructed car insurance form, or where a soldier needs to understand a technical instruction manual ([DuBay, 2004]). It is noted that because the readability formulas attempt to automate results without human input, illogical scores can be attained in applying the equation haphazardly, where for example it is mathematically possible to obtain two identical readability values for texts where the latter for example may contain much more difficult words. In efforts to improve the reliability and accuracy of readability formulas, there are other efforts to quantitatively optimize the text in terms of its readability towards its respective target audience ([Gunning, 1952], [Spandel, 2005], [Halliday, 1976], [William and Thompson, 1988], [Grosz et al., 1995], [DuBay, 2004]). But overall, although the Flesch reading ease test can certainly be incorrectly utilized, when applied properly it is still one of the more popular and easily recognized formulas, used by researchers to this day ([DuBay, 2004],
[Paasche-Orlow et al., 2003], [Williamson and Martin, 2010]).

Looking at the interpretation table in Table 2.1, the ability to read paragraphs of increasing complexity generally increases with age up to the adult level. Accordingly, grade schoolers have a relatively low ability to read complex passages with a score of 80 to 100, and high school students are capable of reading more complex texts with a score of 50 to 60. College educated individuals are capable of reading very complex texts, with a score somewhere between 0 and 30.

The Flesch reading ease formula will be one of the ways in which the texts in the present study are controlled for. In addition to the numerical results provided by the Flesch reading ease formula, the selection of text is itself sourced from grade school, high school, and journal article level writing, which will be detailed in the following sections.

2.2 Methodology

The following sections detail the overall setup of the experiment that is required prior to collecting and analyzing the postural sway data.

2.2.1 Materials: Passage selection and the utilization of SCIgen, and the creation of the final token set

Passage selection for the easy and medium levels

In accordance with the Flesch reading ease readability formula, easy to difficult passages will reflect levels from grade school, high school, and college level scores. Specifically, the spread chosen is 90-95 for "easy", 50-55 for "medium", and 10-15 for "difficult", which also provides a spread of at least 35 points between each level.

Furthermore, the passages are also created from material that should correspond to the targeted scores. In other words, text that should have grade school level scores are chosen from grade school texts, text that should have high school level scores are chosen from high school texts, and text that should have college level scores are chosen from college texts.

For the "easy" and "medium" passages, publicly available and authentic writings by sixth graders and high schoolers are utilized. These writing samples are available at http://www.ode.state.or.us for research use. Authentic representative writing by the target grade level is chosen because it helps to ensure that the vocabulary and grammatical structures do not exceed the particular grade level that is chosen.

For the final selection of the "difficult" level passages, several choices were available for source selection.
In the end, SCIgen, a computer program, was utilized to help create the difficult passages. What SCIgen does is generate unique texts that sufficiently reads like journal articles, in the way that the text is difficult to understand. Further explanation and rationale for choosing SCIgen will be discussed in more detail below.

**Using SCIgen for the difficult level passage selection**

For the difficult passages, because the website that provided the "easy" and "medium" passages at http://www.ode.state.or.us only offers text up to the high school level, an attempt was made to utilize other sources. First, exemplary GRE written samples were tested for usage, but this did not yield an ideal text for the present purposes. In looking over exemplary GRE texts, it becomes apparent that one of the criteria for succeeding in the GRE is the passage length, which would make the text challenging for the wrong reasons, in that for postural sway data collection, the easy, medium, and difficult passages should all be roughly equal in length.

Next, academic journal articles for the "difficult" passages were considered, for example from history articles, but an issue arises involving the fairness of the material. Because the subjects tested are in a university campus, the chance of subjects having knowledge of the text’s contents is present if the content happens to be related to their major. Therefore although good journal article selection helps to alleviate the problem of lengthy writing, some subjects may end up having an unfair advantage over another. For example, if a given test passage deals with history, subjects who happen to be historians will understand the passages much more than subjects who are involved in different majors. If one wishes to minimize the chance of this potential unfairness, at minimum the historians would have to be excluded, but additionally, there could be other subjects who minor in history, took basic history classes, or even look to history as a hobby, further confounding the design of the study.

**SCIgen in the context of text difficulty**

Therefore, the solution chosen was to select writing passages available from SCIgen at https://pdos.csail.mit.edu/archive/scigen/. Ideally, one would like to find a set of difficult passages in which no one has a clear strong foothold and comprehensive advantage over with respect to that particular passage’s content. What this essentially means is a selection of a journal articles focusing on a topic that nobody has a specialty in should be chosen. In ideal terms, this could be argued to be the most fair solution, because it would reflect high quality writing without putting anybody at an unfair disadvantage, as it would be relatively equally hard and challenging for everybody.

Passages from SCIgen are sufficiently difficult, and are written at journal-article level. But what sets
these articles apart from the others is that the writing is randomly generated from an online computer
algorithm. SCIgen itself was an effort to present randomly generated journal articles to reviewers in the
hopes of getting accepted (for which it eventually succeeded in getting accepted, as documented on their
website). Below are a few snippets of such difficult-to-interpret generated texts:

"..Consider the early architecture by A. Gupta et al.; our design is similar, but will actually realize this
objective. We postulate that relational modalities can synthesize consistent hashing without needing to
locate Lamport clocks. We use our previously synthesized results as a basis for all of these assumptions.
Though analysts usually believe the exact opposite, our application depends on this property for correct
behavior..."

"..The deployment of reinforcement learning has been widely studied [32,13,20]. Along these same lines,
the much-touted system by Li [9] does not create DHTs as well as our method [23]. The original method
to this quagmire was well-received; nevertheless, it did not completely solve this question [27]. It remains
to be seen how valuable this research is to the programming languages community. We had our method in
mind before Wilson and White published the recent acclaimed work on erasure coding [22]. The original
solution to this quandary by Martinez et al. [33] was well-received; on the other hand, this discussion did
not completely accomplish this ambition [22]."

Going beyond a shallow reading of the passages generated by SCIgen, fallacies that exist in the text
would naturally reveal themselves to the careful reader. However, subjects in the present study only listen
to each passage once, and their goal from the start is not to determine whether the information is accurate
or not, but rather, to best understand the information presented to them. In a pilot test that included
undergraduates in mechanical engineering, computer science, and electrical engineering, subjects were asked
to listen to the passages and answer a brief quiz on the content, where afterwards, none of the subjects
reported being aware of the fact that the data referred to was computer-generated and did not exist in the
literature. None of the pilot subjects had a sense that the passage material was fabricated by SCIgen (and
this is also true for the actual subjects tested in the experiment as well, who were asked after the experiment
if they had any sense of the difficult texts being illogical or computer-generated).

To summarize, the text for the final token set is firstly constrained by the Flesch reading ease formula,
where three levels of difficulty are created, with scores being 90-95 for "easy", 50-55 for "medium", and
10-15 for "difficult", providing a spread of at least 35 points between each level. In summarizing the passage
selections, the materials for easy and medium difficulty are constructed from grade school and high school
writings. The difficult level is constructed from a journal article level of writing, but more specifically, from SCIgen. SCIgen generates text that while on the surface appears to be from a journal article, in actuality refers to data that does not exist when checked with the actual existing literature. The text generated from SCIgen is chosen for three reasons. From the perspective of creating a uniformly difficult token set, native speakers of English will willingly engage themselves in the SCIgen-generated text if a) they only have the one chance to listen to it and b) are asked to just comprehend the passage for what it is rather than criticize it. Furthermore, because the data referred to in SCIgen text is computer-generated and not taught, c) all the subjects will be exposed to fresh and rather unfamiliar content. Using SCIgen should provide more uniformly difficult passages than when compared to sourcing from real journal articles, in that if real articles were chosen rather than SCIgen, that some subjects majoring in a given discipline might have an unfair educational advantage over other subjects depending on the material that is chosen.

Regarding the construction of the spoken paragraphs, regardless of difficulty level, each paragraph token contains 145 to 149 syllables, and lasts between 30-33 seconds in length. The final token set for the easy, medium, and difficult passages are shown in Table 2.2, Table 2.3, and Table 2.4.

Table 2.2: Tokens for the easy level.

I think Ben’s best trait is his determination. I met Ben through water polo. He plays on the high school team. I’m on the middle school team, but we practice together. Practice starts at four. We get there at three. But before four, we don’t work much at all. We play cards, listen to music, or just talk. On the other hand, Ben is always the first to get to practice. He is always working hard. He comes and runs laps around the track. He only comes back to refill his water bottle. Ben trains the whole time before working just as hard when the real practice starts. To me, that is what real determination is.

Now then, which of the following best represents the overall meaning of the passage?

A) The high school and middle school polo team can practice together in the same place.

B) Nobody from either of the teams has time to play games like frisbee or cards.

C) Real determination to me is to work as hard before and during practice.

D) Any school sport activity can be hard and requires a lot of work.
Tom is my role model. He is also my friend. He is very smart too. Tom has always been in the high level math group. He is also in the highest level reading group. Tom is also really good in the accelerated spelling class too. I don’t know how he could possibly cram all those things in his brain. But he does. He just makes it look so easy, too. Last year, in fifth grade Tom told me he was going to a private school called Green Fields. The next year Tom went to that school, but for just a month. Then he got accepted into Saint Francis Junior High. Only geniuses can attend that school.

Now then, which of the following best represents the overall meaning of the passage?

A) I am smart and I went to a private school only geniuses attend.
B) We wish that Green Fields had a library that we could use to study more.
C) Tom is a role model to me and he is both intelligent and smart.
D) Saint Francis Junior High is a very good school and even better than Green Fields.

Soccer is a challenging sport. You really have to put your mind to it. The one way to become a good player is to just practice and practice. You have to learn how to have a good attitude too. You have to learn how to cheer and encourage all of your teammates. You also have to learn about teamwork. No person should be just a ball hog. And with soccer you can also make more friends too. You will get to know new people, and new people will get to know you. Parents can also get to know each other as well. You could share this fun with your family and friends by inviting them to watch all the people play.

Now then, which of the following best represents the overall meaning of the passage?

A) Cheering the crowd is a big and essential part of any sport in the world.
B) Teamwork has to be learned to play soccer and nobody should ever be a ball hog.
C) The sport of soccer is challenging and is also a great social event.
D) You need the ball at all times to showcase your abilities to everyone else.
Swimming is fun. You get to hold your breath underwater, jump off the diving board, and do cannonballs into the water. You can cool off from the hot sun too, and meet new friends. When you’re at the pool, it is easy to learn new things, like how to doggy paddle, tread water, or do a backstroke, which might save your life one day. You can also learn how to do things like flips and dives. In addition to learning new things, people can burn calories at the pool too. I know someone who lost two pounds from swimming in just three weeks. It is very good to stay fit. Swimming is a fun and great way to do that.

Now then, which of the following best represents the overall meaning of the passage?

A) It is important to learn how to swim because it may one day save your life.
B) Going out to swim makes for a fun hobby and a way to potentially lose weight.
C) When it comes down to it there are no truly good benefits to swimming at all.
D) Swimming is a very challenging experience for the young and old alike.

Anne Smith did not have a good childhood at all. When she was just ten years old, she was sent to a small mental hospital. And later on in her life she was sent to the Pearl Institute for the Blind because her eyes had been hurt from a bad infection. But she tried really hard and still got to graduate from the school. Then one year later she was sent to teach a blind, deaf, and stubborn six year old girl how to spell and basically become a civilized human being. This girl’s name was Jane Doe. Anne did her best and soon she got along with her. She became Jane’s friend and full time teacher for the rest of her life.

Now then, which of the following best represents the overall meaning of the passage?

A) Nobody was able to teach Jane Doe how to become a good and intelligent person.
B) Anne Smith overcame many things in her life and became a good teacher.
C) Jane Doe married a nice young man not long after graduating college.
D) It is possible to have poor eyes yet still graduate from school regardless.
All individuals, no matter what age, want to be accepted in life. Everybody wants to have a place in the sun amongst their peers. Unfortunately, however, some will do whatever it takes to arrive at that coveted destination. Lost Canyon was a recreation spot out in the middle of nowhere, and my friend Alexander and I were vacationing there with a group of other boys. Alexander fixated himself on trying to fit in with everybody else, but when he finally got his place in the sun, he didn’t expect to get into so much trouble.

Now then, which of the following best represents the overall meaning of the passage?

A) A few kids gathered at a particular location that was not very famous.
B) People tend to think logically and rationally when they wish to be accepted.
C) Many decisions in life don’t really require a lot of critical thinking.
D) Alexander was a friend who wanted the spotlight but eventually got in trouble.

I am a varsity jumper for my high school track team and I am proud, but accomplishing this turned out to be more difficult than I anticipated, not only because of the physical demands of the sport, but also from the undue mental stress and pressure exerted by those around me. It is not an untrue rumor that seniors can be excessively intimidating. Even sophomores and juniors can act overly superior, glad to be over the freshman hurdle at last. What can make them really upset is when some silly "fresh" cuts into their time under the spotlight.

Now then, which of the following best represents the overall meaning of the passage?

A) One reason being a freshman jumper is hard is because upperclassmen can give you a hard time.
B) Everybody tends to think that senior athletes have it easy in school.
C) First year athletes are very antagonistic when it comes to sharing the spotlight.
D) Jumping is just a very physically demanding sport to be in.
Today, when walking into an American store or high school you find yourself in the middle of a melting pot of the fashions from the teens of the last thirty five years. Looking around any popular hang-out for teens today, you will see flannel shirts from the grunge period, long, waving skirts of the hippies, oxford cloth shirts, complete with popped collars from the preps, and all black converse hi-tops from the punk predecessors. Granted, there are still people who dress very conservatively, not showing much of a style at all, but, for the most part, the mix of trends has taken over the teen fashion scene.

Now then, which of the following best represents the overall meaning of the passage?

A) It appears that teenagers are not very interested in fashion these days.
B) Less conservative teenagers wear high top sneakers and even oxford cloth shirts.
C) American high school fashion consists of trends made from the past several decades.
D) Fashion today will probably not be that different from fashion in the future.

With modern technology, we have been able to establish relationships with characters at a computer screen even hundreds of miles away. People don’t need to take time out of their lives to create the memories that uphold the foundations of real relationships. The trends of instant messenger and online dating appeal to countless numbers. This is because they are seemingly safe, self-satisfying, immediate, substitute pleasures for real relationships in life, and free of the messy mistakes that come hand in hand. Instant messaging is only just a fad.

Now then, which of the following best represents the overall meaning of the passage?

A) Computers can be used to produce fun relationships and good memories.
B) Building meaningful relationships through the internet works and is here to stay.
C) Relationships through computers look good but can’t replace relationships in real life.
D) All current social media these days have their own unique upsides and downsides.
As teenagers, we are often stereotyped. It’s true that there are many of those who take advantage of the freedoms society gives us. They may also abuse those freedoms. However, I know that there are also many of us who desire to move beyond those misinformed assumptions and are fast becoming dependable, informed members of society. Today I’m proposing that the voting age is conditionally lowered to age sixteen so that these teens can extend their passion and vibrancy to the real world and help shape the society that will someday be theirs.

Now then, which of the following best represents the overall meaning of the passage?

A) High schoolers have a passion and vibrancy that can potentially shape the future.
B) People are all too often busy and ignore the difficulties high schoolers have.
C) Teenagers have never been or will be informed members of society.
D) Teenagers are mature enough to think critically and should be able to vote.

Replicated applications are the norm when it comes to improvement of the location identity split. However, this method is entirely and adamantly opposed. There are many distributed and self-learning methodologies that use trainable information studying congestion control. In this work we allow the UNIDAC converter to construct a reliable technology without visualization of architecture. As a result, we consider how operating systems can apply to extreme programming emulation.

Now then, which of the following best represents the overall meaning of the passage?

A) Methodologies are unreliable and should not be used for extreme programming.
B) Emulation of information carries dense databases and is hard to control.
C) Reliable technologies can be made without visual help or design.
D) A technology is made which has practical implications to programming.
Table 2.4: Continued

Multimodal approaches are quite robust when it comes to virtual information. On this note though, two properties make our method distinct. First, our system studies Gaussian configurations, and also, our methodology visualizes symmetric encryption. But, it should be noted that NO-CEE explores pseudorandom methodologies. NO-CEE is based on the principles of opportunistically crossed software engineering. While similar systems simulate the synthesis of e-business, we answer this issue without developing relational configurations.

Now then, which of the following best represents the overall meaning of the passage?
A) Relational configurations are needed for the main method that is utilized.
B) Multimodal methods are used for simulations in a more independent manner.
C) Specialized encryption and configurations are two unique parts to the method.
D) Many systems will often employ the use of pseudorandom technologies.

Random information and Epsilon fault tolerance garner profound interest from hackers worldwide and electrical engineers in recent years. Distributed applications are particularly compelling when it comes to access points. Existing pseudorandom and perfect methodologies use real time configurations to manage information retrieval systems. The shortcoming of this method, however, is that the lookaside buffer and intercore can synchronize to answer this riddle. The improvement of compilers would improbably degrade compact models.

Now then, which of the following best represents the overall meaning of the passage?
A) Distributed applications could be useful when regarding the access points.
B) Engineers don’t have any interest in random information research.
C) Special information and fault tolerances are affected by compiler tweaks.
D) Information retrieval systems can make useful data for many people.
The understanding of agents has visualized the Internet, and current trends suggest that the evaluation of forward-error correction will soon emerge. To put this in perspective, consider that infamous cryptographers rarely use object-oriented languages to achieve this objective. On a similar note, Without a doubt, we view networking as following a cycle of four phases: management, observation, study, and development. The emulation of semi-randomized algorithms would minimally degrade homogeneous archetypes.

Now then, which of the following best represents the overall meaning of the passage?

A) Forward error correction evaluation will not emerge any time soon.
B) Networking as a repeating cycle generally follows four given phases.
C) Cryptographers don’t really understand the important value errors carry.
D) Current research trends would benefit from the emulation of randomized algorithms.

Telephony and agents, while essential in theory, have not until recently been considered important. Along these same lines, the inability to effect e-voting technology of this result has been resolutely challenged. Furthermore, existing symbiotic and semisymbiotic applications utilize the confirmed unification of fractal nodes and replication to store reinforcement learning. We withhold these algorithms due to space constraints. Unfortunately, model checking alone cannot fulfill the need for the synthesis of compilers.

Now then, which of the following best represents the overall meaning of the passage?

A) Algorithms are not very relevant to any theories of telephony.
B) Model checking alone gives an adequate algorithm to satisfy theories.
C) Reinforcement learning is stored by what are called symbiotic applications.
D) Telephony and agents are very essential to the synthesis of compilers.

Creation of the final token set

Five passages were constructed for each of the three difficulty levels. Each of these passages were controlled for with regards to overall time length being from 30 to 33 seconds long. This time length was achieved
by asking the speaker to read each passage aloud until the length of the audio happened to fall within the
30-33s time range. Furthermore, each passage contains 145 to 149 syllables, with the readability scores
being constrained using the Flesch reading ease formula as described earlier. With five passages and three
difficulty levels, in total fifteen such passages were prepared for the subjects. Audio is recorded in a sound
attenuated booth.

After each passage, a question about the main point of the passage is posed in the form of a 4-choice
multiple choice question. The four choices are constrained in syllable count, containing between 80 and
90 syllables total. The questions are created for the sake of having the subject mentally engaged in the
experiment to do their best.

Each audio file consists of the main paragraph recording, and the four multiple question recordings.
Any recorded speech was re-recorded in the case of any disfluencies as vetted by another native speaker of
English. The comprehension recording itself consists of the statement, "Now then, which of the following
best represents the overall meaning of the passage," which is then followed by the quiz question. Combining
the passage token and the multiple choice questions, subjects stand for about a minute per trial.

The 15 paragraphs are inserted within a total of 64 tokens, where the 49 other tokens are not related
to the passage experiment. The paragraphs are inserted pseudo randomly into every chunk of four other
tokens. This is done because pilot subjects complained about the testing of consecutive paragraph tokens
being draining and fatiguing both mentally and physically.

2.2.2 Procedure

In this study, subjects stand on the force plate while audio is played back to them with their eyes open.
Although closing one's eyes can also make one's balance more sensitive to external perturbations (e.g,
[Davis et al., 2009]), subjects are asked to stand with their eyes open and fixated on a target. When listening
to another person's speech, subjects may arbitrarily tilt their head or look to the side as a matter of habit, and
by instructing them to look at an image while listening, that their sway will be less reflective of unconscious
idle movements such scratching one's nose, tilting one's head, folding one's arms, etc.

Subjects are also instructed to stand without shoes and with their feet in tandem, but otherwise, to
stand naturally. Force plates may provide different results depending on the shoe ([Hatton et al., 2013]).
In standing in tandem, the dominant foot is placed directly in front of the nondominant foot (Figure 2.1).
When people stand, they rely on a sufficiently adequate base of support to maintain their stability (e. g.,
[Whiting and Rugg, 2006]). Reducing the base of support by having subjects stand in tandem decreases this
stability ([Bandy and Sanders, 2007]), as a decrease in the base of support makes subjects more sensitive
to external perturbations ([Lowes et al., 2004]). After they listen to the token, a multiple choice question is played back to them. Upon hearing the entire question, they orally provide the correct answer "a", "b", "c", or "d". Postural sway is not recorded during the questioning, as the questions are created for the sake of keeping the subject mentally engaged in the experiment to do their best.

Additionally, five trials without any speech data are randomly inserted into the experiment. These trials serve the purpose of keeping the subject alert during the trials. After each such trial, the subjects are then asked if they wish to take a break and rest on a chair. There are 64 trials in total, with twelve trials being relevant to the present study, and the rest serving as fillers.

Figure 2.1: Standing in the tandem position. The dominant foot is placed directly in front of the nondominant foot.
2.2.3 Subjects

Data from 25 male native speakers of English (age 23.4+/−5.32 years, weighing 77.5+/−15.9kg, height 180+/−8.74cm) are analyzed in the present study. One of the subjects reported being left-hand and left-foot dominant, with the remaining twenty four subjects being right-handed and right-foot dominant. Subjects who reported any vision or hearing problems, or any presence of neurological or balance disorders were excluded from the study, and likewise, any subjects that had considerable trouble maintaining the tandem stance having to resort to a non-tandem stance mid-experiment were excluded. No subjects reported any exercise in their weekly regimen that focuses heavily on balance training, such as taichi or yoga. All subjects are 18 years or older, and this study has approval from the Institutional Review Board.

2.3 Data Analysis

2.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial distance data over time

We used R (R Core Team, 2012) and lme4 (Bates, Maechler and Bolker, 2012) to perform a linear mixed effects analysis of the relationship between difficulty and the fiftieth percentile of the radial range of the COP. The dependent variable is the fiftieth percentile of the radial range of the COP. As fixed effects, we entered difficulty level into the model, with three levels. As random effects, we had intercepts for subjects. The distribution of the resulting fiftieth percentile of the radial range of the COP values from the subjects’ sway data tended to be right-skewed, so a log transformation was taken prior to submission for statistical analysis. P-values were obtained by using the mixed command, where a significant effect is found for level (F(2,46)=4.62, p=.01). Post-hoc tests reveal a difference between easy and difficult (F(1,50)=4.82) and medium and difficult levels (F(1,50)=8.52). Box plots for the data across difficulty level are shown (Figure 2.2).

Because we are also interested in the different trends of the postural sway over time for each token, we also model the radial sway data over time with GAMs. The dependent variable is the radial distance of the COP. Spline fitted data is generated by the GAM for each of the tested paragraphs (Figure 2.4). Fixed effects are taken as difficulty level with three levels. Subject is taken as a random intercept, and token is taken as a random slope. Overall, 28 percent of the deviance in all the sway data is explained by the GAM, which falls within the range of other studies using GAMs (e.g. [Víkingsson et al., 2016] and [Zanobetti et al., 2000]).
The Effect of Paragraph Difficulty Level on Postural Sway

Figure 2.2: Plot showing that the log of the fiftieth percentile radial range of the COP is higher for the difficult level in relation to the medium and easy levels.
Figure 2.3: Demonstrative figure representing the trend of the radial distance data in the paragraph tokens, where the x axis is the time axis, and the y axis is the radial distance. One token is chosen per difficulty level, which shows the general observed trend across difficulty level. Difficulty level is color coded such that blue is easy, yellow is medium, and red is difficult (for black and white printing, darker crisscross patterns indicate the harder difficulty level). The region of interest is the entire duration of the token. Over each difficulty level, a general trend is observed, where with regards to the overall slope of the radial distance over time, easy tokens are downwards-sloping, difficult tokens are flat-sloped, and medium tokens are somewhere in between. The more difficult the token is, the greater the slope of the radial distance becomes. For more detailed plots of all the tokens, see Figure 2.4.
Figure 2.4: A series of time plots that show the effect of paragraph difficulty on the slope of the radial distance of the COP. The plots are the result of fitted splines from the GAMs of the radial distance of the COP versus time, for each of the fifteen tokens. Difficulty is color coded from blue to red from the top row to the bottom row (for black and white printing, from light to dark crisscrosses). Easy passages are in the first row, medium passages in the second row, and difficult passages in the third row.

2.3.2 Discussion of the observed trends between postural sway and paragraph difficulty level

The results from the mixed models support the general hypothesis that more difficult passages incur larger amounts of postural sway. There is a significant effect of level, where the difficult passages have higher values of the fiftieth percentile of the radial range of the COP than the medium and easy passages. In other words, difficult passages tend to elicit larger differences in postural sway as represented by the fiftieth percentile of the radial range of the COP. This result is intuitive in the sense that more difficult passages elicit greater changes in overall balance.

For the GAMs, the percentage of deviance explained for the paragraph tokens is sufficiently high enough such that interpretations of the trends in the plots can be made. The percentage of deviance explained for the GAMs on the paragraph tokens is at 28 percent, for which when utilizing GAMs, serves a similar purpose as to how adjusted R squared values are used. In other words, similar to the adjusted R squared, the percentage of deviance explained helps to understand how much of the postural sway data can be explained by the GAM. In a way that’s also similar to the adjusted R squared value, the percentage of the deviance explained has an ideal value of 100 percent, but in practice, a wide range of values comes out in actual
research, where for example values as different as 15 percent and 70 percent have been reported in the literature. The output of the GAM in the present study falls within the range of the percentage of deviance explained from other literature, where useful interpretations of the trends of the data from the GAMs can be made.

Similar to the results from the analysis on the fiftieth percentile of the radial range of the COP, an intuitive interpretation of the results from the GAM is also present in observing the radial distance data over time. Demonstrative plots are chosen to illustrate the trend, shown in Figure 2.3. One token is displayed for each of the three difficulty levels, and the region of interest in each of the three plots is the entire duration of the token. In the easy level, the radial distance of the COP tends to slope downwards. In contrast to the easy tokens, the overall slope of the radial distance for the medium level is more positive, in the sense that it is a less negative slope. For the difficult tokens, likewise, the overall slope is even more positive than both the easy and medium passage tokens. In other words, if the passages are easy, then the subjects' radial distance of the COP will tend to approach a minimum over time, but if the subjects encounter more difficult passages, the time taken to approach this minimum takes longer to achieve. The full set of plots are shown at Figure 2.4.

If one thinks of more difficult passages as eliciting greater agitation or mental concentration that reflects in their balance, the results make intuitive sense. For easier passages, subjects settle down in their sway relatively quickly, compared to the medium and difficult passages. For the medium passages, it takes longer than in the easy passages for the subjects to eventually settle down. Likewise, for the most difficult passages, it takes even longer compared to the medium passages to approach a minimum, implying perhaps that compared to the easy and medium passages, that they are in a consistent state of agitation or concentration throughout the entirety of the passage that affects their postural sway.

2.3.3 Summary of the paragraphs experiment

This chapter established the link between postural sway and the difficulty level of paragraphs. Overall, sway is affected by the content of the paragraphs for difficulty level. Paragraphs that are more difficult tend to have a larger fiftieth percentile of the radial range of the COP, and within the tokens, difficulty can alter the sway with regards to the radial distance over time by offsetting the time at which a minimum of the radial distance is approached, which has different trends for each difficulty level. These results support the hypothesis that difficult language makes for increased changes in postural sway, and it supports the notion that language can perturb standing balance.
Chapter 3

Changes in Balance Corresponding to the Aural Processing of Single Sentences

3.1 Introduction

In this chapter we experiment on sentences. It was learned in the previous chapter that the difficulty level of paragraphs can influence postural sway. However, in sentences, some more care will likely be needed because the words used are much fewer than paragraphs. Therefore to make allowance for the natural variation of words but in a controlled database of sentences, word frequency and syllable structure are controlled for, with a focus maintained on the effects of grammatical structure on postural sway during sentence processing.

3.1.1 Background: Choosing tightly controlled sentence pairs

Controlling for both word and sentence structure at the same time is not trivial but necessary, and one way in which one can control for both is to utilize sentence pairs that only differ in their word order. Listening to sentences can invoke changes in cognitive load ([Tettamanti et al., 2005], [Michael et al., 2001], [Humphries et al., 2001], [Just et al., 2008]), incurring different levels of cognitive load as a consequence of that particular construction ([Frazier and Fodor, 1978], [Frazier, 1998], [Marslen-Wilson and Tyler, 1978], [McClelland and Rumelhart, 1981], [Friederici, 2002]). If sentence types are taken that only differ in their word order, one can invoke differing levels of cognitive load based on the ordering of the words alone. One could take these sentence pairs and observe how these sentences would affect changes in cognitive load as reflected through postural sway.

In this chapter, the sentences utilized for the stimuli are heavy noun phrase shift (HNPS) sentences and subject-subject and subject-object (SS/SO) relative clauses, which will both be detailed in the section below.

3.1.2 HNPS and SS/SO Relative Clauses

To create a balanced stimuli set involving postural sway for sentences, inherent variation of individual words is accounted for in that sentence pairs that only differ in word order are chosen. For example, constructions such as (1a) and (1b),
are fairly different constructions, even though they use the same words.

To better understand this particular sentence pair, the latter construction (1b) is considered awkward or ungrammatical compared to the former ([Kimball, 1973], [Stallings and MacDonald, 1998]), and this phenomenon relates to the switched positions of the bracketed noun phrases. A noun phrase (NP) refers to the word or words that function as a subject, object, or prepositional object in the sentence. Sentences of the latter construction type in (1b) are generally less acceptable unless there is something "heavy", ie longer, in the bracketed NP.

The bracketed NP, "a fork", can occur in two positions in English. The sentence in (1a) is an acceptable sentence with regards to its word order. But in (1b), the bracketed NP is shifted to the right. The sentence in (1b) is considered awkward, but the heavier, ie the longer, the NP is, the more likely and the more acceptable it is for the NP to be shifted to the final position. This phenomenon is referred to as the heavy noun phrase shift (HNPS). Accordingly, in (2a) and (2b),

(2a) "Jack gave to Lisa [a brand new yet thoroughly expensive gold and silver plated fork]"
(2b) "Jack gave to Lisa [a fork]"

(2a) is considered much more acceptable than (2b), because the NP is heavy in (2a), and not heavy in (2b).

Therefore in taking advantage of the fact that when the NP is non-heavy, that a sentence with a shifted NP is considered less acceptable than its non-shifted counterpart, potential changes in cognitive load can be observed from HNPS pairs such as (1a) and (1b) as the differences may be reflected through postural sway data.

Another sentence construction type that can incur differing amounts of cognitive load by just changing the word order are in (3a) and (3b):

(3a) "The man who quickly shot the boy killed Bill"
(3b) "The man who the boy quickly shot killed Bill"

Taking a closer look at the sentences, both contain a relative clause that modifies the subject of the sentence. However, (3a) and (3b) differ in the role the main subject noun phrase ("the man") plays in the relative clause. In sentences such as in (3a), the main-clause subject is also the subject of the verb in the relative clause (such that the man shoots the boy). In sentences such as in (3b), the main-clause subject is
the object of the verb in the relative clause (such that the boy shoots the man). These sentence pairs are known as Subject-subject (SS) and Subject-object (SO) relative clause sentences.

Both SS and SO sentences are equally unambiguous, but it has been shown with regards to comprehension that the latter construction (3b) requires greater levels of cognitive load ([Miller and Chomsky, 1963]), such that a larger magnitude of electrical activity is present in the brain when processing these SO sentences ([Muller et al., 1997]). The exact theoretical mechanism for why SO sentences may incur more cognitive load is a point of debate ([Muller et al., 1997]), but regardless of whether the reasoning is because of working memory ([Miller and Chomsky, 1963]), or differences in grammatical processing ([Wanner, 1978]), the literature is supportive of the fact that the difference in cognitive load required for processing sentences such as (3a) and (3b) is present ([Tavakolian, 1981], [King and Just, 1991], [Cooke et al., 2003], [Grossman et al., 2003]).

Therefore, in taking advantage of the fact that the SO variant of the sentence is established as being more cognitively demanding than the SS variant, potential changes in cognitive load can be observed from sentences like (3a) and (3b) as the differences may be reflected through postural sway data.

To sum, both of these sentence types mentioned above are established in the previous literature as having a difference in how they are accepted by native speakers, and for present purposes, should help alleviate the problem of word variability when testing only for sentence effects. For HNPS sentence types, native speakers of English avoid any word ordering with the direct object placed at the end of the sentence if the noun phrase is not heavy ([Stallings et al., 1998]), and likewise, native speakers require less processing demands for SS sentences compared to their SO counterparts ([Miller and Chomsky, 1963]). Therefore both of these sentence types offer a way to test the viability of using force plates to measure cognitive load in using the dual-task paradigm.

In an attempt to keep the token set in the present experiment tightly controlled for with regards to words, syllable count and word frequency are also controlled for, which will be detailed further in the next section.

3.2 Methodology

The following sections detail the overall setup of the experiment that is required prior to collecting and analyzing the postural sway data.
3.2.1 Materials: Creating a token set controlled for with respect to possible word effects

General design of the token set

Six HNPS sentences (Table 3.1) and six SS/SO sentences (Table 3.2) are created for use in the present study. This set of sentences was split in two, such that subjects are exposed to one of two token sets, where each has no overlapping alternative form. In other words, to prevent a given subject from being repeatedly exposed to similar sentence content, they were only presented with one of the two possible word orderings for each of the tokens. As a result, each subject was exposed to three tokens each for the two possible word orderings, with six sentences in total.

Controlling for possible word effects

Outside of word order, an effort is also made in keeping the respective sentence pairs as similar as possible. In other words, because the grammatical differences are what’s important in the present study, only words that are frequently occur are used. This is done by checking with various pooled databases containing word frequencies in the English language, choosing only common words are chosen, such as "moon", or "map". (The occurring frequencies of these words ranged from 717 to 24 words per million words, which allows for fairly common words to be used. For a more detailed procedure regarding the selection of the words as it relates to this frequency range, see the methodology section in chapter 5. Words falling under the easiest difficulty levels of I and II were chosen, which falls under the range of words that occur fairly frequently in the English language.)

Additionally, in controlling for the database on sentences, if a token requires names, no rare names are used, as any names used in the token set are chosen from the 500 most popular list of names in America from 1990-2000 as according to the social security registry. Moreover, the HNPS sentences contain exactly six words and six syllables, and SS/SO sentences contain exactly nine words and twelve syllables. For both sentence types, any word that was more than one syllable long has the stress placed on the same syllable for every other token. For example, if the second word of the token set was a two-syllable word with the stress on the first syllable, then all the other tokens' second word contains a two-syllable word with the stress on the first syllable.
Table 3.1: HNPS token set, where the easier sentences are on the top set of rows, and the harder sentences on the bottom set of rows.

<table>
<thead>
<tr>
<th>Less difficult</th>
<th>More difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ross made a map for Neil.</td>
<td>Ross made for Neil a map.</td>
</tr>
<tr>
<td>Sean caught a fish for Paul.</td>
<td>Sean caught for Paul a fish.</td>
</tr>
<tr>
<td>Todd bought a watch for Jack.</td>
<td>Todd bought for Jack a watch.</td>
</tr>
<tr>
<td>Keith paid a bill for Rose.</td>
<td>Keith paid for Rose a bill.</td>
</tr>
</tbody>
</table>

Table 3.2: SS/SO token set, where the easier sentences are on the top set of rows (SS sentences), and the harder sentences on the bottom set of rows (SO sentences).

<table>
<thead>
<tr>
<th>Less difficult</th>
<th>More difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>The lawyer who called the senate lost the package.</td>
<td>The lawyer who the senate called lost the package.</td>
</tr>
<tr>
<td>The father who left the daughter watched the marriage.</td>
<td>The father who the daughter left watched the marriage.</td>
</tr>
<tr>
<td>The teacher who helped the writer loved the college.</td>
<td>The teacher who the writer helped loved the college.</td>
</tr>
<tr>
<td>The leader who sent the mayor held the hearing.</td>
<td>The leader who the mayor sent held the hearing.</td>
</tr>
<tr>
<td>The female who killed the model told the nation.</td>
<td>The female who the model killed told the nation.</td>
</tr>
<tr>
<td>The mother who found the ladies sold the knowledge.</td>
<td>The mother who the ladies found sold the knowledge.</td>
</tr>
</tbody>
</table>
Construction of the audio files used for testing

The sentence tokens were read aloud by a native speaker of English (male, age 35), which was recorded in a sound attenuated booth. Any recorded speech was re-recorded in the case of any disfluencies as vetted by another native speaker of English.

Each audio file consists of the main sentence recording, and a true-false comprehension question recording. The comprehension question consists of "True or false? According to the statement.." which is then followed by a quiz question. The syllable and word counts for each question is controlled, with HNPS true-false questions containing seven words and seven syllables, and SS/SO true-false questions containing five words and seven syllables. The duration of the entire trial is about 7s for the HNPS sentences, and 8s for the SS/SO sentences. The choice of which token set to use for each subject is chosen at random each time.

3.2.2 Procedure

In this study, subjects stand on the force plate while audio is played back to them. They are instructed to stand in the tandem position, have their hands at their side, and look at an image placed in front of them. Subjects are told to stand naturally and not make deliberate movements such as adjusting one's glasses, or crossing one's arms. After they listen to a sentence, a randomly chosen true-false question from the database of tightly controlled questions is played back to them. Upon hearing both the sentence and the question, they orally provide the answer "true", or "false".

Additionally, five trials without any speech data are randomly inserted into the experiment. These trials serve the purpose of keeping the subject alert during the trials. After each such trial, the subjects are then asked if they wish to take a break and rest on a chair. There are 64 trials in total, with twelve trials being relevant to the present study, and the rest serving as fillers.

3.2.3 Subjects

Data from 25 male native speakers of English (age 23.4+/−5.32 years, weighing 77.5+/−15.9kg, height 180+/−8.74cm) are analyzed in the present study. One of the subjects reported being left-hand and left-foot dominant, with the remaining twenty four subjects being right-handed and right-foot dominant. Subjects who reported any vision or hearing problems, or any presence of neurological or balance disorders were excluded from the study, and likewise, any subjects that had considerable trouble maintaining the tandem stance having to resort to a non-tandem stance mid-experiment were excluded. No subjects reported any exercise in their weekly regimen that focuses heavily on balance training, such as taichi or yoga. All subjects are 18 years or older, and this study has approval from the Institutional Review Board.
3.3 Data Analysis

3.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial distance data over time

We used R (R Core Team, 2012) and lme4 (Bates, Maechler and Bolker, 2012) to perform a linear mixed effects analysis of the relationship between difficulty, sentence type, and the fiftieth percentile of the radial range of the COP. The dependent variable is the fiftieth percentile of the radial range of the COP. As fixed effects, we entered sentence type and difficulty into the model, each with two levels. As random effects, we had intercepts for subjects and token set. The distribution of the resulting fiftieth percentile of the radial range of the COP values from the subjects’ sway data tended to be right-skewed, so a log transformation was taken prior to submission for statistical analysis. P-values were obtained by using the mixed command, where although no significant effect was found for sentence type (F(1,69)=2.28, p=.14) or interaction of sentence type and difficulty (F(1,69)=.01, p=.92), a significant effect is found for difficulty (F(1,69)=4.35, p=.04). Box plots for the fiftieth percentile of the radial range of the COP across difficulty are shown (Figure 3.1).

Because we are also interested in the different trends of the postural sway over time for each token, we also model the radial sway data over time with GAMs. The dependent variable is the radial distance of the COP. Spline fitted data is generated by the GAM for each of the tested sentences (Figure 3.4 and Figure 3.5). Fixed effects are taken as sentence type and difficulty, each with two levels. Subject and token set are taken as a random intercept, and token is taken as a random slope. Overall, 23.2 percent of the deviance in all the sway data is explained by the GAM, which falls within the range of other studies using GAMs (e.g. [Víkingsson et al., 2016] and [Zanobetti et al., 2000]).

3.4 Discussion

The following sections are devoted to help in understanding the results of the present study on sentences.

3.4.1 Discussion of the observed trends between postural sway and sentences

The results from the mixed models support the general hypothesis that difficult sentences incur larger amounts of postural sway. There is a significant effect of difficulty, where the more difficult variant of the sentence type is larger in value. In other words, difficult sentences tend to elicit larger amounts of changes in postural sway as represented by the fiftieth percentile of the radial range of the COP.
Figure 3.1: Both the HNPS and SS/SO sentence types have a relatively easy and difficult variant, from which the difference comes from only switching the word order. There are therefore a total of four sentence variants as shown in the plot, two of them being relatively easy, and two of them being relatively difficult compared to its respective counterpart. Each of the sentence type’s difficult variants are highlighted in a black crisscross pattern, while the relatively easier sentence variants are in clear white.
Figure 3.2: Demonstrative figure representing the trend of the radial distance data in the sentence tokens, where the x axis is the time axis, and the y axis is the radial distance. This token pair shows the general observed trend over the token’s relative difficulty level. The only difference between a token pair is the word ordering. The easy token is on the left, and the hard token is on the right. The colored region indicates the region in time where the word order between the two tokens differ, such that the blue region is for the easier token, and the red color is for the hard region (for black and white printing, the harder region is shaded in a crisscross pattern). The dark black vertical line at the end of the colored region indicates where the tokens’ respective difference in word order has finished. The region of interest has been circled, which is the region of time after the colored region. After the colored region, the more difficult variant (plot on the right) has a higher uptick in sway than the easier variant (plot on the left). For more detailed plots of all the tokens, see Figure 3.4.
Figure 3.3: Demonstrative figure representing the trend of the radial distance data in the sentence tokens, where the x axis is the time axis, and the y axis is the radial distance. This token pair shows the general observed trend over the token’s relative difficulty level. The only difference between a token pair is the word ordering. The easy token is on the left, and the hard token is on the right. The colored region indicates the region in time where the word order between the two tokens differ, such that the blue region is for the easier token, and the red color is for the hard region (for black and white printing, the harder region is shaded in a crisscross pattern). The dark black vertical line at the end of the colored region indicates where the tokens’ respective difference in word order has finished. The region of interest has been circled, which is the region of time after the colored region. After the colored region, the more difficult variant (plot on the right) has a higher uptick in sway than the easier variant (plot on the left). For more detailed plots of all the tokens, see Figure 3.5.
Figure 3.4: Resulting fitted splines from the GAMs of radial distance versus time, for the HNPS tokens. In this figure, only the word orders are different between the token pairs in each row. These plots are titled as unshifted (plots in the left column) and shifted (plots in the right column), where shifted has incorrectly structured word order, and unshifted preserves the more natural word order. The region of time where the word orders differ from each other are highlighted in different colors, where the easier token (unshifted) is shaded in blue, and the more difficult token (shifted) shaded in red. The thick vertical black bar represents the end of the colored region, i.e., when the difference between the two token pairs’ word order is respectively completed.
Figure 3.5: Resulting fitted splines from the GAMs of radial distance versus time, for the SS/SO tokens. In this figure, only the word orders are different between the token pairs in each row. The region of time where the word orders differ from each other are highlighted in different colors, where the easier token (SS) is shaded in blue, and the harder token (SO) shaded in red. The thick vertical black bar represents the end of the colored region, ie when the difference between the two token pairs’ word order is respectively completed.
For the GAMs, the percentage of deviance explained for the sentence tokens is sufficiently high enough such that interpretations of the trends can be made. For the results on the GAMs, the percentage of deviance explained for the GAMs on the sentence tokens is at 23.2 percent, for which when utilizing GAMs, serves a similar purpose as to how adjusted R squared values are used. In other words, similar to the adjusted R squared, the percentage of deviance explained helps to understand how much of the postural sway data can be explained by the GAM model. In a way that’s also similar to the adjusted R squared value, the percentage of the deviance explained has an ideal value of 100 percent, but in practice, a wide range of values come out in actual research, where for example values as different as 15 percent and 70 percent have been reported in the literature. Considering that that the percentage of the deviance explained in the present study falls within the range of the percentage of deviance explained from other literature, this percentage of deviance explained in the present study appears adequate such that meaningful interpretations on the trends of the radial range data over time can be made.

Similar to the results from the analysis on the fiftieth percentile of the radial range of the COP, an intuitive interpretation of the GAM results is present with the information gleaned from the radial distance data over time. Demonstrative plots are chosen to illustrate the observed trend, shown in Figure 3.2 and Figure 3.3. Each figure has plots of one token pair with the easy and difficult variant. Recall that each token pair uses the same set of words. The parts where the order of the words differ per token pair are delineated in the colored regions, where the more difficult variant is in red, and its easier counterpart in blue. The end of the region is denoted by a thick black vertical bar. After the subject finishes listening to the words in the colored regions, there is a higher uptick in sway for the difficult token than for the easier variant. In other words, after the subject has finished processing the set of words that should have been more difficult to process as denoted by the colored regions, the slope of the radial distance is relatively higher compared to its easier counterpart.

This trend is true for 5 of the 6 tokens for the HNPS sentences, and 4 of the 6 tokens for the SS/SO sentences. The full set of plots for both sentence types are in Figure 3.4 and Figure 3.5.

Put into physical meaning, after the critical parts of a sentence are processed, the radial distance of the COP is destabilized depending on the relative difficulty of the sentence. Along the region denoted by the thick black vertical line where the words in the colored regions have been fully processed, the slope for the harder tokens is generally higher than for its easier counterpart, and this is true for the majority of the HNPS and the SS/SO sentences. Subjects tend to destabilize in the form of greater postural sway when there are elements in the sentence that are harder to process.
3.5 Summary of the sentences experiment

This chapter established the link between postural sway and the relative processing difficulty of sentences. Overall, sway is affected by the content of the sentence. Sentences that are perceived to be more difficult tend to have a larger fiftieth percentile of the radial range of the COP, and in looking at the radial distance of the COP as a function of time, difficulty can destabilize sway by increasing the relative slope of the radial distance of the COP at the difficult regions. These results generally support the hypothesis that difficult language makes for increased changes in postural sway, where language can effectively perturb standing balance.
Chapter 4

Changes in Balance in a Replication Study on Acceptability Judgments for Sentences

4.1 Introduction

This chapter comprises the second half of the study on sentences and postural sway, expanding on postural sway and sentences in the form of a replication work. In other words, a separate focus is now placed on sentences and balance, this time in the way of experiment replication. In this part of the sentence studies, sentence data will be utilized from another researcher’s work. An attempt is made to replicate their findings, with balance from force plates instead of the ranking methods they had used. A successful replication of their overall results would further demonstrate the efficacy of balance changes in the way they can tap into changes in cognitive load, while simultaneously validating the other researcher’s overall methodology too.

The present chapter looks at balance data while replicating a study by [Juzek, 2016] on sentence acceptability, with four of the different sentence types. The acceptability data from that study are then compared with postural sway data attained from the present study.

4.1.1 Background: Replication and verification of a study on the acceptability of sentences

The material that will be used for an attempt at replication is in a thesis by [Juzek, 2016]. In the second chapter of his thesis, he looks at acceptability judgments of various types of sentences, where subjects processed a number of grammatically correct and incorrect sentences under different modes of presentation. In other words, when humans judge the acceptability of a sentence, the author was interested in how people rate the acceptability of that sentence when presented in a formal or casual user visual interface, combined with the sentences presented in a written or spoken manner.

Subjects participated in the experiment online using Amazon’s Mechanical Turk, which provided the overall visual and audio framework for his study. In his computer survey, he varied the formality and mode of presentation, such that the sentences could be presented to a subject with a formal or informal user interface, and with the material presented as either spoken or written. No significant differences were found
from observing the subjects’ responses between the different modes of presentation and formality. The results suggest that regardless of context in which the language material is presented, that acceptability rankings are fairly robust.

Because a) the study deals with acceptability judgments, where b) the sentence token data is publicly available, and c) the numerical rankings are also reported in his work, four of the author’s utilized sentence types are used in the present study, as these four are constructed with a relatively greater focus on listening comprehension. These constructions are the alt-if constructions, resumptive pronouns, good sentences, and bad sentences. As in previous chapters, these sentences will be presented auditorally to the listener, and will be detailed further in the methodology section.

Returning to the goals of the present chapter, one of the main goals in this chapter is to reproduce similar findings from the work by [Juzek, 2016]. In any research field, the replication and verification of an experiment is an important component of quality research ([Hempel, 1968], [Hempel and Oppenheim, 1948], [Lakatos and Musgrave, 1970], [Meehl, 1990], [Platt, 1964], [Salmon and Earman, 1999]). In various sciences, there are works that try and place their efforts on replicating important experimental works, with varying degrees of success ([Collaboration, 2015]). If the replication is successful, it verifies the theoretical science behind the actual protocol, and if not, it leads to further discussion about the targeted material ([Ioannidis, 2005], [Begley and Ellis, 2012], [Prinz and Asadullah, 2011], [McNutt, 2014], [Pashler and Wagenmakers, 2012]). However, as important as it may be, reduplication in and of itself does not tend to be very high with regards to overall research priority ([Nosek and Lakens, 2014], [Schmidt, 2009]), which provides one driving incentive for the work in this present chapter.

The second main goal of the present study relates back to the central theme of the present thesis, which is to perform the verification study by measuring standing balance. If one could verify sentence acceptability ratings in the work by [Juzek, 2016] by using force plates in the present work, then the two mentioned goals can be tackled at the same time. Such a replication experiment would allow one to verify that force plates can detect changes in cognitive load, as well as verifying the methodology as applied in the work by [Juzek, 2016]. [Juzek, 2016] makes an experiment based off of acceptability and introspection, and takes advantage of modern technology by using Amazon’s Mechanical Turk to obtain an estimate for the acceptability of various sentence constructions. Although the goal of the author’s experiment has more to do with the effects of the mode of presentation on grammatical acceptability, the author presents the results of the data with sufficient details which can be used in a replication study using force plates and measuring postural sway.

There are scientific reasons behind this reasoning. There is a chance that the results of an experiment
for a statistical result are erroneous due to biased sampling, sparse data, and overfitting, and/or reasons leading to type I or type II errors. If the experimental results are replicated using different subjects in a different lab, then the validity of the experiment result is a lot stronger. Additionally, if different methodologies are employed to examine a phenomena from different viewpoints and the same result is obtained, it further strengthens the probability that the observed/measured phenomenon is correct, i.e., the chance of the alternative interpretation that both studies are subject to the same type I or type II error is greatly reduced.

To summarize, tokens used in [Juzek, 2016] will be used in the present chapter. In the present chapter it is investigated whether balance can be used to verify the works performed in a study of acceptability judgments. The present study collects balance data using four sentence types from [Juzek, 2016] in a replication study to evaluate whether responses from two drastically different methodologies, native subjects’ evaluation via Mechanical Turk, and postural sway measured using a force plate, show similar trends.

4.2 Methodology

The following sections detail the overall setup of the experiment that is required prior to collecting and analyzing the postural sway data.

4.2.1 Materials used from the study on acceptability

Of the tokens created by [Juzek, 2016], four sentence constructions are particularly relevant to the current study. These sentences are good sentences, alternative if-clauses, resumptive pronouns, and bad sentences, where all four sentence types were taken from spoken sources (Table 4.1). Although there are more sentence types tested in his work, only the stimuli which were designed for spoken language and presented in an auditory format are used.

Description of the token set

The four chosen sentence types from [Juzek, 2016] are detailed as follows:

First, (a) alternative if-clauses (e.g., "If he would do X, he would do Y") are sentences often documented to appear in spoken language ([Miller et al., 1998]). These sentences are grammatical and acceptable in the English language.

Second, (b) sentences with resumptive pronouns used in [Juzek, 2016] involve a personal pronoun in a relative clause restating the word, phrase, clause, etc., after a pause or interruption (e.g., "this is the man that I don’t know where he lives"). These sentences are not generally viewed as being very acceptable. However,
Table 4.1: Sentence tokens for the replication study.

<table>
<thead>
<tr>
<th>Alt-If</th>
</tr>
</thead>
<tbody>
<tr>
<td>If she would come to see things for herself, she would change her mind.</td>
</tr>
<tr>
<td>If you would go on such a journey, you would understand what he is talking about.</td>
</tr>
<tr>
<td>If we would spend every day connecting with people, we could learn more about the world, too.</td>
</tr>
<tr>
<td>If I would go to New York myself, I might get the chance to meet Lady Gaga.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resumptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is a donkey that I don’t know where it lives.</td>
</tr>
<tr>
<td>This is the man that I don’t know where he comes from.</td>
</tr>
<tr>
<td>You fear things that you don’t understand what they are.</td>
</tr>
<tr>
<td>We are afraid of things that we don’t know what they are.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is our job to figure out what happened and do everything we can to prevent it from happening again.</td>
</tr>
<tr>
<td>Steven Colbert always threatens to run for office in South Carolina, but of course, always with a wink.</td>
</tr>
<tr>
<td>You can also make the argument that they didn’t approve anything.</td>
</tr>
<tr>
<td>And in his absence, the Senate Republicans pushed through a redistricting plan for the State Senate.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>If was key word ‘hope’ in 2009, perhaps ‘change’ key word 2013.</td>
</tr>
<tr>
<td>1986 the last time was really it happened.</td>
</tr>
<tr>
<td>I feel was really thrown to wind caution these times and really far more aggressive he were.</td>
</tr>
<tr>
<td>First term president Obama start out with numbers of envoy I think.</td>
</tr>
</tbody>
</table>

...they are sometimes used to self-correct the on-the-fly flow of information when speaking mid-sentence, as they do in fact occur in spoken language ([Prince, 1990]).

Third, (c) good sentences are sentences taken from spoken news sources (e.g., "And in his absence, the Senate Republicans pushed through a redistricting plan for the State Senate"). The author only alters the tokens in cases of the sentences being too long, and otherwise are all grammatical acceptable sentences.

Fourth, (d) bad sentences are sentences originally taken from spoken sources, but deliberately mangled by [Juzek, 2016], with a focus on incorrectly and intentionally altering word order, agreement, and omission of function words (e.g., "I feel was really thrown to wind caution these times and really far more aggressive he were").

**Recording of the token set**

Each sentence type includes four individual tokens as shown in Table 4.1. The sentence tokens were read aloud by a native speaker of English (male, age 35), and were recorded in a sound attenuated booth. Recorded speech was vetted by another native speaker of English and sentences with any disfluencies was re-recorded.
4.2.2 Procedure

In this study, subjects stand on the force plate while audio is played back to them. They are instructed to stand in the tandem position, have their hands at their side, and look at an image placed in front of them. Additionally, subjects are told to stand naturally and not make deliberate movements such as adjusting one's glasses, or crossing one's arms, and after they listen to a sentence, a true-false comprehension question is played back to them. Upon hearing both the sentence and the question, they orally provide the correct answer "true", or "false".

Additionally, five trials without any speech data are randomly inserted into the experiment. These trials serve the purpose of keeping the subject alert during the trials. After each such trial, the subjects are then asked if they wish to take a break and rest on a chair. There are 64 trials in total, with twelve trials being relevant to the present study, and the rest serving as fillers.

Token syllable and word count were not controlled for, but the syllable and word counts for the questions afterwards are, with each true-false comprehension question containing ten words and twenty syllables. Combined in total, each token varies in time but has around ten to fourteen seconds of spoken data for which the subject must stand. There are 64 trials in total, with sixteen trials being relevant to the present study, and the rest serving as fillers.

4.2.3 Subjects

Data from 25 male native speakers of English (age 23.4+/−5.32 years, weighing 77.5+/−15.9kg, height 180+/−8.74cm) are analyzed in the present study. One of the subjects reported being left-hand and left-foot dominant, with the remaining twenty four subjects being right-handed and right-foot dominant. Subjects who reported any vision or hearing problems, or any presence of neurological or balance disorders were excluded from the study, and likewise, any subjects that had considerable trouble maintaining the tandem stance and having to resort to a non-tandem stance mid-experiment were excluded. No subjects reported any exercise in their weekly regimen that focuses heavily on balance training, such as taichi or yoga. All subjects are 18 years or older, and this study has approval from the Institutional Review Board.
4.3 Data Analysis

4.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial distance data over time

We used R (R Core Team, 2012) and lme4 (Bates, Maechler and Bolker, 2012) to perform a linear mixed effects analysis of the relationship between sentences and the fiftieth percentile of the radial range of the COP. The dependent variable is the fiftieth percentile of the radial range of the COP. As fixed effects, we entered sentence type into the model, with four levels. As random effects, we had intercepts for subjects. The distribution of the resulting fiftieth percentile of the radial range of the COP values from the subjects' sway data tended to be right-skewed, so a log transformation was taken prior to submission for statistical analysis. P-values were obtained by using the mixed command in lme4, where no significant effects are found ($F(3,14.35)=.28, p=.84$). Box plots for the data across sentence type are shown (Figure 4.1).

Because we are also interested in the different trends of the postural sway over time for each token, we also model the radial sway data over time with GAMs. The dependent variable is the radial distance of the COP. Spline fitted data is generated by the GAM for each of the tested sentence tokens (Figure 4.6). Fixed effects are taken as sentence type with four levels. Subject is taken as a random effect, and token is taken as a random slope. Overall, 21.6 percent of the deviance in all the sway data is explained by the GAM, which falls within the range of other studies using GAMs (e.g. [Víkingsson et al., 2016] and [Zanobetti et al., 2000]).

4.3.2 Discussion of the observed trends between postural sway and sentences

Because the results from the mixed models are somewhat different than the rest of the experiments so far, GAMs will be discussed first. Information about the general trends of sway can be gleaned from the radial distance data over time, where spline fitted data is generated by the GAM for sentences. Each of the four sentence types follow their own respective patterns.

For the GAMs, the percentage of deviance explained for the sentence tokens is sufficiently high enough such that interpretations of the trends can be made. For the results on the GAMs, the percentage of deviance explained for the GAMs on the sentence tokens is at 21.6 percent, for which when utilizing GAMs, serves a similar purpose as to how adjusted $R$ squared values are used. In other words, similar to the adjusted $R$ squared, the percentage of deviance explained helps to understand how much of the postural sway data can be explained by the GAM model. In a way that’s also similar to the adjusted $R$ squared value, the percentage of the deviance explained has an ideal value of 100 percent, but in practice, a wide range of values come out in actual research, where for example values as different as 15 percent and 70 percent have been
Figure 4.1: The tokens from four general sentence types were taken from a study by [Juzek, 2016], with the resulting plot of sway versus each of the four sentence types. The less acceptable sentence types are crisscrossed in the plot, and the more acceptable sentence types are in clear white.

Figure 4.2: Demonstrative figure representing the general trend observed for the alt-if tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. It is observed that the radial distance generally follows a downwards sloping trend over the entire token. For more detailed plots of all the tokens, see Figure 4.6.
Figure 4.3: Demonstrative figure representing the general trend observed for the resumptive tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. The region in time where the sentence can no longer become salvageable in terms of its grammaticality is highlighted in red. (For black and white printing, the red region is also shaded in a crisscross pattern.) When the subject listens to the sentence in this less acceptable region (circled above), there is a relative uptick in sway. For more detailed plots of all the tokens, see Figure 4.6.

Figure 4.4: Demonstrative figure representing the general trend observed for the good tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. It is observed that the radial distance generally follow a downwards sloping trend over the entire token. For more detailed plots of all the tokens, see Figure 4.6.
Figure 4.5: Demonstrative figure representing the general trend observed for the bad tokens. The x axis is time, and the y axis is the radial distance. One token is chosen to illustrate the overall trend. It is observed that the radial distance fluctuates more relative to the trends in the other sentence types. For more detailed plots of all the tokens, see Figure 4.6.

reported in the literature. Considering that that the percentage of the deviance explained in the present study falls within the range of the percentage of deviance explained from other literature, this percentage of deviance explained in the present study appears adequate such that meaningful interpretations on the trends of the radial range data over time can be made.

The following figures that are referenced below are demonstrative plots to show the overall trend in the data.

The trend for the alt-if sentences, shown in Figure 4.2, tends to follow a general downwards sloping trajectory of the radial distance of the COP over time across its tokens. Nothing is grammatically incorrect about these particular tokens, so this observed trend makes sense in the way one might intuitively expect, in that the trend is relatively consistent across each of its tokens. All these tokens were designed to follow a fairly similar sentence structure (If i would do X, she would do Y, etc.), so for the most part, it makes sense that the four alt-if tokens’ radial distance data behave roughly similarly. The tokens are designed with a degree of consistency that more or less reflects itself in the trends of the radial distance data in the form of a generally downwards-sloping trend.

The trend for the resumptive tokens is shown in Figure 4.3, and the plot is split into red (crisscrossed for black and white printing) and blue colored regions. The blue region represents the beginning of the sentence, for example, "This is the man that I don’t know" in the sentence "This is the man that I don’t know where
Postural sway tokens reflecting sentence acceptability

**Alt-if tokens**

**Resumptive tokens**

**Good tokens**

**Bad tokens**

Figure 4.6: Resulting fitted splines from the GAMs of radial distance versus time, for the tokens in the verification experiment. Parts where the sentence is expected to be easy are highlighted in blue, and hard highlighted in red. For black and white printing, the red region is also highlighted in dark crisscrosses.
he lives." The region colored in blue is acceptable until the region in red is reached. The region where some difficulty might be encountered in the sentence is highlighted in the figure in red, ie "This is the man where I don’t know [where he lives]", where the bracketed region in the sentence corresponds to the region of the figure shaded in red and crisscrossed.

The radial distance data in the resumptive tokens tend to follow a downwards or neutrally sloping trend until about halfway through each trial, where there is an uptick in sway present for the majority of the tokens. The location of the change occurs when the sentence starts to become less acceptable and more difficult to follow (corresponding to the highlighted red regions in the figure). In other words, the overall trend from the GAMs makes sense in generally reflecting the fact that the sentences start out as grammatically correct but end up being less acceptable later on. The trends in the resumptive token set in the start of the sentences are generally consistent in its being either downwards or neutral-sloped, where for a majority of the tokens, the region where the sentence begins to become difficult as denoted by the red-shaded region has a higher slope when compared to the acceptable portion of the sentence before it. In other words, the subjects’ postural sway response to the sentences are generally perturbed in the regions where difficulty might be expected across each of the tokens. The trend of the radial distance data behaves as one might intuitively expect, with the relative slope of the radial distance increasing in the regions where difficulty might be present.

Relative to the other tokens tested, the trend for the bad sentences in Figure 4.5 do not have a consistent upwards or downwards pattern, which makes sense with respect to its original experimental design, in that the only criterion for designing these sentences were that the sentences were not meant to be acceptable, but not controlling for the locations where the sentences start being confusing. There is variation across the tokens because there is variation in the sentence structure and word content, ie the sentences are just designed to be not acceptable, and the content and word order of these sentences were not controlled for, which logically can lead to the within-token variation as shown. In other words, the results suggest that radial distance data can vary by a lot between tokens, such that they can behave differently depending on its internal content.

Lastly, the observed trend for the good sentences in Figure 4.4 has a generally downwards-sloping radial distance data over time, and this consistency makes sense when compared to the trends of the bad token sentence set. Both token sets are similar in the way they were designed, in that grammatical structure and word content are not controlled for, which can create a lot of variation in each sentence’s structure. However, unlike the bad sentence types, subjects should have no real difficulty in processing acceptable sentences in general. This difference is shown in the form of a more consistent trend being present in the good token set when compared to the bad token set. Even though the word content and grammatical structure are not
controlled for in either token set, because the sentences are all acceptable in the good token set and are fairly
easy to process for the subjects, not too many differences are present in the trends across the good sentence
tokens, at least in relative comparison to the variation that is present in the bad token set.

It is also noted that the general downwards trend in the good token set is also present for the alt-if token
set. The similarity between the two token sets make sense in the way that both are acceptable sentence types
to native speakers of English. Additionally, the variation within the alt-if tokens appear more diminished
internally than the token-to-token variation within the good token set. This could also be attributed to the
fact that the alt-if tokens follow a much more rigid sentence structure (If i would do X, she would do Y,
etc.), while the good token set does not have this particular factor controlled for.

The comprehensive plots with all the tokens for each of the sentence types as referenced above can be
seen in Figure 4.6.

Overall, it seems that the radial distance of the COP can show consistent trends if the designed tokens
are either acceptable sentences, or less acceptable but consistent in grammatical structure. For example,
the good token set and the alt-if token set both show a generally consistent downwards trend, which can
be explained by the fact that well-formed sentences by themselves would not perturb the balance of the
subjects too much. Likewise, the majority of the resumptive tokens show a consistent trend with regards to
having a larger slope in the regions in red, implying that consistent results can be had if the sentence is less
acceptable but designed with a consistent grammatical pattern. The relatively large amount of variation of
the radial distance data in the bad token set implies that if no general grammatical structure is followed,
the results of the radial distance data will likewise vary accordingly if the sentence is not very acceptable.

Moving onwards, the results from the mixed models do not support the general hypothesis that more
difficult sentences incur larger amounts of postural sway. The final boxplot of the data suggests no observable
trend of the sway, and difficult sentences don’t elicit larger amounts of changes in postural sway as represented
by the fiftieth percentile of the radial range of the COP. However, it is reasonable to assume that sentence
length would affect the postural sway. A follow-up study will be discussed in more detail in the following
section.

4.3.3 Factoring in syllable count

It is reasonable to expect that the range of variation in syllable count will not affect sentence acceptability
judgments in the study by [Juzek, 2016], as listeners or readers make the judgment after the sentence is
completed. However, variation in syllable count will impact an on-line, time-sensitive measurement such as
postural sway. Syllable count was not taken into account when designing the present study, and it may be
adversely affecting the data results. The results may make more sense if one factors in the syllable count. The present section will evaluate data when syllable count is normalized.

At first look, it appears that the fiftieth percentile of the radial range of the COP is not a good indicator of acceptability judgments given the lack of significant results in the previous section. However, in thinking about possible confounds to the present work, there is one noted difference between this chapter and the previous chapter on sentences. This difference is with regards to controlling for the syllable count of each token. Therefore, the focus is now placed on the potential confounding factor of syllable count, and how it may relate to postural sway.

The token set in the study by [Juzek, 2016] is loosely controlled for at the word or syllable length within token types, though not for tokens across sentence types, as shown in Table 4.2. While such a design doesn’t affect sentence acceptability judgments, it may affect postural sway across token types in the current study.

### 4.3.4 A preliminary plot factoring in syllable count

One difference between the studies so far is that the tokens used in the present study differ in the control of the syllable count. Therefore, it’s possible that the results could make more sense if another look is taken at the data this time accounting for syllable count. As a preliminary analysis, a graph is made accounting for syllable count to get a visual idea of this possibility (Figure 4.7) using the following methods.

Inherent subject variation had previously been accounted for by using the mixed models approach and taking subjects as a random intercept. In this graph, to likewise accommodate for inherent subject variation, every subject’s sway data is normalized by the median of his own sway values. In other words, to make the data better comparable across subjects, the fiftieth percentile of the radial range of the COP data for each subject is normalized by the median of his fiftieth percentile of the radial range of the COP. For all the trials for each subject, the median value of fiftieth percentile of the radial range for the COP is determined, and then for each of his trials, each radial range value is divided by this median value. This effectively creates a range of values centered about a value of 1. Converting this to percent, each person’s radial range value is centered at 100%.

Using this subject normalization procedure, as a preliminary investigation to understand how syllable count might impact postural sway results, the results from each sentence type is simply divided by its respective average syllable count, and the results of the plot are shown in Figure 4.7. Although this result is just a preliminary result, it does suggest that syllable count might be an important component to consider when using postural sway as a means to measure changes in cognitive load. In the figure, the sentences that are expected to be less acceptable have higher amounts of postural sway, and the sentences that are more
The Effect of Sentence Acceptability on Postural Sway

Figure 4.7: Plot of the normalized fiftieth percentile of the radial range of the COP for the sentence types in the verification study, this time normalizing by the syllable count. The tokens from four general sentence types were taken from a study by [Juzek, 2016], with the resulting plot of sway versus each of the four sentence types. Postural sway data is normalized by subject and divided by the average syllable count of each of the tokens per sentence type. The less acceptable sentence types are crisscrossed in the plot, and the more acceptable sentence types are in clear white.
Table 4.2: Tokens used from the replicated study has differing syllable counts across sentence types.

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Avg. syllable count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alt-If</td>
<td>20.75</td>
</tr>
<tr>
<td>Resumptive</td>
<td>12.25</td>
</tr>
<tr>
<td>Good</td>
<td>22</td>
</tr>
<tr>
<td>Bad</td>
<td>18.5</td>
</tr>
</tbody>
</table>

acceptable have lower amounts of postural sway.

4.3.5 A more formal analysis factoring in syllable count

For the purpose of the present work, it is plausible to consider that a time-related dependence on the tokens exists as it relates to postural sway. More specifically, there may be a relationship between postural sway and syllable count that can and should be investigated. With the data from his study that is available, an effort is now made to mathematically analyze the kind of relationship syllable count may have on postural sway.

An OLS regression is used to investigate the possible relationship between acceptability and postural sway. The OLS regression takes the form of the following equation:

\[ y = \varepsilon + \beta_1 x_1 + \beta_2 x_2 \]

\( y, x_1, \) and \( x_2 \) are known and inputted into the above equation, such that the goal is to solve for the unknown coefficients \( \varepsilon, \beta_1, \) and \( \beta_2. \) In this equation, I will assume that sway is dependent on acceptability and speaking rate, where speaking rate is defined as the total number of syllables for a token divided by the total time taken for the utterance. Therefore, \( y \) is sway, \( x_1 \) is acceptability, and \( x_2 \) is speaking rate.

However, in thinking about the coefficients in the OLS regression, it is likely that the relationship between physical body movements and a pencil-and-paper numbered scale will not be linear. Therefore, rather than assume a linear relationship, a power relationship is assumed.

Additionally, in order to intuitively understand the relative weights of the coefficients, the sway, acceptability, and speaking rate are normalized. Each of the subjects’ data points are divided by his median for that parameter, such that the resulting data is centered at 1.

Therefore, the equation for such a relationship would look like

\[ \text{NormalizedSway} = \varepsilon \ast (\text{NormalizedAcceptability})^{\beta_1} \ast (\text{NormalizedSpeakingRate})^{\beta_2} \]

which does not follow the original form of the OLS regression, but may perhaps more accurately model the relationship between sway and acceptability.
To solve for the coefficients, the above equation is first manipulated so that it more closely follows the original form of the OLS regression equation. The equation is transformed in such a manner by taking the natural log of both sides:

\[
\ln(\text{NormalizedSway}) = \ln(\varepsilon) + \beta_1 \ln(\text{NormalizedAcceptability}) + \beta_2 \ln(\text{NormalizedSpeakingRate})
\]

Solving for the coefficients yields values of \(\varepsilon = 1.03\), \(\beta_1 = -0.103\), and \(\beta_2 = 0.320\). Of particular importance are the signs of \(\beta_1\) and \(\beta_2\), because \(\beta_1\) relates to the score of acceptability, and \(\beta_2\) relates to the speaking rate of the speaker. From an intuitive standpoint, if the sentence is more acceptable, postural sway should decrease. Likewise, if the speaker speaks at a faster rate, postural sway should increase. Observing the signs of the coefficients, coefficients \(\beta_1\) and \(\beta_2\) are negative and positive, respectively, which preserves this intuition.

The OLS regression allows us to convert the acceptability scores into postural sway values. Plotting the theoretical values versus the actual sway values, one might expect the slope of the best fit line to be close to 1, and the R squared value to be close to 1. The slope and R squared value is .946 and .977 respectively, further suggesting that syllable count plays an important role in sway studies when dealing with acceptability.

To sum, an OLS regression was taken to better understand the relationship between postural sway and acceptability. Assuming that postural sway depends on both acceptability and speaking rate, the sway, acceptability, and speaking rate data is normalized and the equation is adjusted such that a power relationship instead of a linear relationship is solved for. Importantly, the signs of the coefficients match what one might intuitively expect, such that more acceptable sentences reduce sway, and faster speech increases sway. Additionally, the resulting slope and R squared value of the best fit line for the converted versus actual values of postural sway is close to 1, suggesting that a relationship between sway and acceptability can be found if syllable count is accounted for.

### 4.3.6 Summary of the verification study

To conclude, the findings from trends of the radial distance of the COP make intuitive sense in the way that subjects’ postural sway is destabilized in the regions in time that are relatively less acceptable in the given sentence. In other words, subjects are sensitive to the acceptability of sentences, and it shows in the form of increased postural sway at the regions of time where difficulty may occur.

With regards to the fiftieth percentile of the radial range of the COP, although the initial findings of postural sway with respect to the replication of previous data seem to have suggested contrary findings, accounting for the syllable count allows for a more understandable interpretation of the data. In the studies of aural comprehension in the previous chapters of this thesis, syllable count was very tightly controlled.
Figure 4.8: When syllable count is factored into the regression equation, the transformed sway data closely matches the numbers from the acceptability judgments with a slope very close to 1 at 0.946. The R squared value is also at .977, which further suggests that postural sway may have a relationship with syllable count.
for. In this replication study, however, syllable count is not. An ordinary linear regression is taken on the data not to find the relationship between sway and acceptability, but rather, to better understand the relationship between sway and syllable count. Considering that the resulting R squared value is greater than .9, it is plausible that sway seems to be related to approximately the square of the averaged syllable count ratio. Further work will benefit in understanding this tentative relationship for postural sway and syllable count, and with this knowledge one can better find a relationship between sway and syllable count with more accuracy in future studies.

Overall, from the present and previous chapters on language and balance there is evidence supportive of the fact that both paragraphs and sentences can affect postural sway. Therefore last but not least, in the following chapter a focus is placed on vocabulary words, observing its overall effect on postural sway. In thinking about language involved with the communication of just single words, one final question in the present thesis is posed: What is the effect of processing vocabulary words on postural sway?
Chapter 5

Changes in Balance Corresponding to the Aural Processing of Vocabulary Words

5.1 Introduction: Choosing words under a set of controlled guidelines

In this final study, instead of sentences and paragraphs, the primary goal is to determine whether postural sway can be used to gauge how difficult aurally presented vocabulary words are. More specifically, an effort is made to understand the relationship between balance data and the listening comprehension of word difficulty for native speakers of English. The overall prediction is that the subjects will sway more when the word difficulty is increased.

It is noted that vocabulary words may be too short to exert an influence on changes in postural sway. For instance, postural sway literature has recommended 60 seconds of data for reliable results ([Doyle et al., 2007]), and spoken words by themselves would be considerably shorter in duration than complete sentences, not to mention paragraphs. Therefore, the next several sections will detail how the tokens were designed to address this issue. Paragraphs, which contain sentences and vocabulary words, are the largest linguistic element for which postural sway has been tested on. Vocabulary words, by contrast, can be considered as the smallest element. This study on vocabulary words helps to understand how small of a linguistic element postural sway can be used for, with regards to providing an intuitive interpretation of changes in balance as it relates to changes in cognitive load.

Moving on to the general criteria of the token creation, there are numerous factors worth considering. For instance, words in general may differ not only in their syllable count, but also in their syllable structure. Furthermore, for the purposes of the present study, these words should somehow range from an easy to a difficult level. In an attempt to consider the prevalent variations that may occur in words, the final set of vocabulary words for testing were chosen with the following requirements in mind:

1-The words contain exactly two syllables
2-The words should not be emotionally charged (e.g., terror, racist, nazi, etc.)
3. The words should follow five types of vowel-consonant patterns
4. The words should have the stress on the first half
5. The words should match one of seven word frequency ranges averaged from two corpus sites
6. The words should not contain different meanings with the same pronunciation (e.g., tear and tier are pronounced the same but have different meanings)
7. One set of words should contain alliteration with an identical CV pattern as one of the other four CV patterns
8. With all these constraints, there should be enough words to fill all seven levels and all five vowel consonant patterns (totaling 35 words)
9. The words should be intuitively difficult or easy for its respective level
10. A set of nonsense words should be created to keep subjects from falling into a lull and claiming that they know every word without seriously thinking about the word. (These words follow rules 1 through 9 wherever applicable.)

The implementations of the ten bullet points are detailed more carefully in the following section, starting with word frequencies and collections of language data known as corpus data, and also a phenomena related to word distribution, known as Zipf’s Law, which will also be detailed in the subsections below.

5.2 Methodology

5.2.1 Materials: Creating the word token set by utilizing word frequency and Zipf’s Law

Creating the vocabulary token set

For the present study, two-syllable is selected as the word length of all stimuli in the experiment, and these words should somehow range from an easy to a difficult level. Regardless of the objective method chosen to rank the words by difficulty for this experiment, the easy words on the easy levels should be intuitively easy, and hard words on the hard levels should be intuitively difficult. Therefore, one possibility would be to take advantage of the fact that more difficult words tend to occur less frequently.

For example, "moon" is a relatively common word, and may be considered as less difficult than the word "hammock", which occurs less frequently in everyday language. Because word difficulty has been correlated with word frequency ([Breland, 1996]), word frequency is used as a guideline for attaining easy to difficult words. To determine the final word set, a four step process is performed as described below, utilizing the
data provided from various institutions that have recorded such frequency information (referred to, hereafter, as corpus data).

**Step 1: Finding shared words between corpuses**

First, an objective method is needed in order to select a range of word frequencies. Frequency data is chosen from three different corpus websites, as these three websites provide the top 5,000 used words in American English. They are from word list frequencies from articles such as Newsweek and local newspapers (http://www.wordfrequency.info/free.asp?s=y), word list frequencies from tv and movies (https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/TV/2006/1-1000), and word frequencies based off of internet search engine results (http://www.insightin.com/esl/). In step 1, each list of the 5,000 words is taken and quartered into four frequency bands of 1,250 words apiece. Each word that occupies all three lists and is in the first band is considered a word that is in "level 1" difficulty. Likewise, any word in all three lists and in the second band number is considered as "level 2" difficulty, etc. These words are not the final list, but serve as an intermediate step, as a set of words that can give an idea of how frequent or infrequent a word has to be in order to be classified at a specified level of difficulty. This list consists of words that all share the same level and also exist in all three corpus data.

**Step 2: Calculating the medians per level**

In the second step, one representative frequency for every band is calculated. First, an averaged word frequency for the relevant word list is calculated. Word frequency is calculated by finding out how many times a target word has been mentioned in a corpus data set, and then dividing that count by the total number of words present in the corpus, for each corpus of interest. For each and every two syllable word in the four-level list, two spoken word frequency measures are obtained that can be manually obtained on a word per word basis in the following two websites: Orally spoken data from the Michigan Corpus of Academic Spoken English (http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple), and orally spoken data from the Corpus of Contemporary American English (http://www.wordfrequency.info/free.asp?s=y). This gives four columns of data for spoken frequency that pertain to four initial levels of word frequency. To get a representative value of word frequency for each of the four levels, the median of each of the four columns is taken. The median is taken and not the average, because very common words have a very high frequency value, which can skew the frequency results, especially for the first column which is filled with very common words. In number of occurrences per million of words, the resulting target frequency values are: Level one: 191.6, level two: 28.5, level three: 15.2, and level four: 10.4.
Step 3: Utilizing a linear domain with Zipf’s law

In this particular step, frequency max/min ranges are calculated instead of frequency values, and those frequency ranges are labeled as the difficulty levels. Ideally, if testing four levels for example, every word chosen in the final list would have the aforementioned values of exactly 191.6, 28.5, 15.2, 10.4, and the word list would be easily determined. However, getting words with that exact degree of frequency accuracy would require full and unadulterated access to the raw data for both the Michigan Corpus of Academic Spoken English and the Corpus of Contemporary American English, neither of which are available to the public. Moreover, the chance of a vocabulary word having the exact targeted frequencies might not be high in and of itself, considering that only two-syllable words are to be used. Therefore, words that fall within a certain tolerance range of the target values are required. If the four median points were equally spaced apart, one could create a simple tolerance range of a constant value for each data point, for example, a spread of +/-10 points per median value. But the tricky part comes from the fact that the four data points follow more of an exponentially decreasing trend than it is a linear one (Figure 5.1). This goes hand in hand with the fact that more common words have much higher and more variable values, whereas less frequent words vary less in magnitude. So, median values in the easy levels should have a larger tolerance range than values in the harder levels. In corpus data for example, the difference between a hard word and a very hard word would only differ in frequency of perhaps 5 points, whereas an easy word versus a very easy word could vary on the scale of hundreds.

To resolve this tolerance issue, a uniform range is used to compute the corresponding frequencies. This frequency range data can then be used for the final word selection process. Moreover, it is known that word frequency follows Zipf’s law, which can be utilized for the present purposes. Zipf’s law is an observed phenomenon in natural language, where the log of the frequency of the word follows a linear relationship to the log of its given rank ([Zipf, 1935], [Zipf, 1949]). Therefore, looking at levels 1, 2, 3, and 4, and selecting one tolerance range for the levels, such as a given percentage away from a given level, one can calculate the corresponding frequency values associated with the given rank. In other words, the frequency values relate to the values of the levels in that the log-log plot of the data can be fitted with a straight line (Figure 5.1). In finding the equation for this straight line, one can calculate the corresponding frequency values for any given level range instead of just the original four levels.

50 percent of the distance from each level for the frequency tolerance range is chosen, which gives frequency values at the levels 0.5-1.5, 1.5-2.5, 2.5-3.5 etc. This leeway is large in the sense that it does not adequately separate the medians from each other. However, one can still sufficiently accommodate for this leeway, by selecting vocabulary words from every other level instead of every single level. In other words,
Step 4: Making difficult vocabulary sets, and constraining syllable structures

The final vocabulary word list consists of vocabulary words with two syllables, and only vocabulary words with stress on the first syllable are chosen (e.g., biker, shovel, not hotel, success, etc.). Taking the average of the two frequency values from the audio corpus data from the Michigan Corpus of Academic Spoken English and the Corpus of Contemporary American English, words are manually searched for that fit within the appropriate tolerance levels. As it turns out, the most difficult level has a frequency of about 8 words per million, but vocabulary words found at this frequency level are not considered to be very difficult. So to make more difficult levels, the linear equation provided from Zipf’s law is utilized, going beyond the fourth level, and up to levels that have frequencies less than one word per million. The final chosen levels are 1,
3, 5, 7, 9, 13, and 21. The average frequency for the hardest level, level 21, is at about 0.26, and these vocabulary words should be sufficiently difficult for the present purposes.

In one more final attempt to control the data, each of the vocabulary words were constricted to follow a given particular syllable structure. The syllable structures used are CVCV, CVCVC, CVCVC (alliterated), VCCVC, and CCVCVC. In this notation, "C" stands for consonant, and "V" stands for vowel. So for example, a two-syllable word that follows a CVCV structure is "lady", and a two-syllable word that follows a CCVCVC structure is "planet". Although the structure CVCVC and the structure CVCVC (alliterated) appear identical on the surface, the latter structure is filled with "alliterated" vocabulary words. These words start with the same consonant for each syllable, for example, "puppet", or "mammal". Because these CVCVC vocabulary words alliterate, they have a simpler structure than CVCVC vocabulary words that don’t alliterate, which has a different starting consonant for each syllable. It may be possible that less cognitive load may be incurred for alliterated vocabulary words when compared to the corresponding set of non-alliterated vocabulary words. The final vocabulary word list consists of five vocabulary words per level, and a total of five levels (Table 5.1).

Lastly, a set of nonsense words is created that obey each of the aforementioned rules. Because listeners will be tested on their vocabulary knowledge, nonsense words are created in order to prevent listeners from claiming that every word is a word without really thinking about it.

Audio stimuli

Audio is recorded from two native speakers of English (male, aged 33 and 35) and are instructed to read aloud the final vocabulary word token set. Any recorded speech was re-recorded in the case of any disfluencies as vetted by another native speaker of English. Each participant listens to a mixture of both speakers randomly assigned for each token.

Although each of the chosen nouns have a strict two syllable requirement, the length of time in each token can vary. In making sure that the subjects are investing their attention in a similar way for all the speech files, the audio data are of equal time length. In order to control for the time length, each word is uttered six times. Postural sway data from just one utterance may be too little data for postural sway analysis, given that at least 60 seconds of postural sway data is recommended for reliable results ([Doyle et al., 2007]). Using a computer code, each of three identical utterances are spaced apart equally within a 3 second time window, followed by a brief pause, and then a final repeat of the same utterance three more times. This gives both sufficient listening data for the listener to process, and it gives sufficient time for acquisition of balance data. There is a one second pause between each set of utterances. In total, postural sway is recorded
Table 5.1: The final selection of the vocabulary words chosen, based on frequency levels. The table shows the vocabulary words chosen; associated frequency data is noted below each of the respective tokens, in words per million. Level is listed in the first column, and the respective upper and lower frequency tolerance ranges are listed below it. Nonsense words are listed in the last row, and have no associated frequency data. In column 4, the abbreviation "allit" next to the syllable pattern stands for vocabulary words containing alliteration.
for seven seconds for each vocabulary word.

5.2.2 Subjects

Eighteen male native speakers of English (age $21+/\pm 8$ years, weighing $80+/\pm 17$kg, height $168+/\pm 45$cm) are recruited in the present study. Two of the subjects reported being left-hand and left-foot dominant, with the remaining sixteen subjects being right-handed and right-foot dominant. Subjects did not report any vision or hearing problems, nor any presence of neurological or balance disorders. No subjects reported any exercise in their weekly regimen that focuses heavily on balance training, such as taichi or yoga. All subjects are 18 years or older, and this study has approval from the Institutional Review Board.

5.2.3 Experimental procedure

For every trial, once the audio starts playing, postural sway is recorded. Subjects are told beforehand that they will be listening to vocabulary words, where they have to either explain the word or repeat the word if they know it. They are instructed to listen to each of the speech files while standing on the force plate, with their eyes open and looking at an image. In this study, subjects stand on the force plate while audio is played back to them. They are instructed to stand in the tandem position, have their hands at their side, and look at an image placed in front of them. Additionally, subjects are told to stand naturally and not make deliberate movements such as adjusting one’s glasses, or crossing one’s arms.

If the subjects are trying to understand the audio passage, the hope is that when they listen to each of the speech samples, the greater the difficulty, the greater a disturbance they will have in their postural sway. After each audio file is played back, subjects may be asked whether or not they know the word, and if they do, to repeat it back to the researcher ("Do you know this word? If you do, please repeat it."). To ensure that the subjects are paying attention, there are several nonsense words (e.g., "plobore", "orvock") incorporated into the data set. Subjects are told about this beforehand, so they know they have to pay attention to the words in order to correctly answer the question. There are five nonsense words, and forty words in total, so 12.5% of the data set consists of nonsense words. For one randomly chosen word in each column in the table of words, instead of being asked to repeat the word, they are asked to explain what the word means ("Do you know this word? If you do, explain what it means."). Lastly, five of the trials are just quiet, ie no words are presented and no questions are asked, and are randomly placed in the experiment as an additional attempt to keep the subject alert and aware during the trials.

The order in which the vocabulary words are presented are random. Subjects are also instructed to fill out a general questionnaire form. In total, testing takes about 40 minutes per subject, with about 30
minutes devoted to measuring postural sway. Because several studies suggest that there may be an effect of gender as it pertains to balance ([Yoshida et al., 1982], [Panzer et al., 1995], [Kim et al., 2010]), data is collected from only male native speakers of English.

5.3 Data analysis

5.3.1 Results from the fiftieth percentile of the radial range of the COP and the radial distance data over time

We used R (R Core Team, 2012) and lme4 (Bates, Maechler and Bolker, 2012) to perform a linear mixed effects analysis of the relationship between difficulty and the fiftieth percentile of the radial range of the COP. The dependent variable is the fiftieth percentile of the radial range of the COP. All seven seconds of all six word utterances are taken when calculating the dependent variable. As fixed effects, we entered difficulty level and syllable type into the model, with seven and five levels, respectively. As random effects, we had intercepts for subjects and speaker. The distribution of the resulting fiftieth percentile of the radial range of the COP values from the subjects’ sway data tended to be right-skewed, so a log transformation was taken prior to submission for statistical analysis. P-values were obtained by using the mixed command, where a significant effect is found for level (F(6,577.33)=2.51, p=.02), and a marginally significant effect for syllable type (F(4,577.45)=2.35, p=.05), and no effect of interaction of syllable type and difficulty (F(24, 576.4)=1.15, p=.28). For difficulty level, post-hoc tests reveal a difference between levels 1 and 13 (F(5,577)=4.49) and 7 and 13 (F(5,577)=4.43). Box plots for the data across syllable (Figure 5.2) and level (Figure 5.3) are shown.

Because we are also interested in the different trends of the postural sway over time for each token, we also model the radial sway data over time with GAMs. The dependent variable is the radial distance of the COP. Spline fitted data is generated by the GAM for each of the tested vocabulary words (Figure 5.5). Fixed effects are taken as difficulty level and syllable type, with seven and five levels, respectively. Subject and speaker are taken as a random intercept, and token is taken as a random slope. Overall, 21.3 percent of the deviance in all the sway data is explained by the GAM, which falls within the range of other studies using GAMs (e.g. [Víkingsson et al., 2016] and [Zanobetti et al., 2000]).
Figure 5.2: Plot showing the effect of word difficulty on postural sway. More difficult levels are shaded in a darker crisscross pattern, such that the more difficult the level, the darker the crisscross pattern. Words were chosen from various difficulty levels based off of the assumption that more difficult words occur less frequently. Words on level 1 are fairly easy (ie commonly occurring), and words on level 21 are more difficult (ie less frequently occurring).
Figure 5.3: Plot showing the effect of syllable structure on postural sway. More complex syllable structures are shaded in a darker crisscross pattern. "C" stands for consonant, and "V" stands for vowel, such that a word like "toe" for example has a structure of CV, "cat" has a structure of CVC, etc. Note that CVCVC (allit) is the alliterated version of the CVCVC syllable structure type.
Demonstrative postural sway pattern for vocabulary word tokens

Figure 5.4: Demonstrative figure representing the general trend observed for the vocabulary words. The x axis is time, and the y axis is the radial distance. One plot is chosen for each of the seven difficulty levels to illustrate the trend. The plots are arranged from top to bottom in order of increasing difficulty level. In looking at the seven plots, it is observed that as the difficulty level increases, the overall slope of the radial distance becomes more negative. For more detailed plots of all the tokens, see Figure 5.5.
Figure 5.5: Resulting spline fits from the GAMs for the radial distance of the postural sway over time. Each subplot represents a particular token. The syllable type is displayed in the columns (first column is CVCV, second column is CVCVC, third column is CVCVC (alliterated), fourth column is VCCVC, and fifth column is CCVCVC). The levels are displayed in the rows (first row is level 1, second row is level 3... last row is level 21). The levels are also shaded from blue to red, or in black and white printing from lighter to darker crisscross patterns).
5.3.2 Discussion of the relationship between postural sway and vocabulary words

The next sections discuss the findings from the results on the processing of vocabulary words as it relates to changes in postural sway.

The relationship between postural sway and vocabulary word difficulty and syllable structure

The results from the mixed models support the general hypothesis that more difficult vocabulary words incur larger amounts of postural sway. While a marginally significant effect is found for syllables, there is a significant effect of level, between levels 1 and 13, and 7 and 13. The final box plot of the data across level suggests a general increase in the amount of sway as difficulty level increases. In other words, difficult vocabulary words tend to elicit larger amounts of changes in postural sway as represented by the fiftieth percentile of the radial range.

Observing the trends from the GAMs

For the GAMs, the percentage of deviance explained for the vocabulary word tokens is sufficiently high such that interpretations of the trends can be made. For the results on the GAMs, the percentage of deviance explained for the GAMs on the vocabulary word tokens is at 21.3 percent, for which when utilizing GAMs, serves a similar purpose as to how adjusted R squared values are used. In other words, similar to the adjusted R squared, the percentage of deviance explained helps to understand how much of the postural sway data can be explained by the GAM model. In a way that’s also similar to the adjusted R squared value, the percentage of the deviance explained has an ideal value of 100 percent, but in practice, a wide range of values come out in actual research, where for example values as different as 15 percent and 70 percent have been reported in the literature. Considering that that the percentage of the deviance explained in the present study falls within the range of the percentage of deviance explained from other literature, this percentage of deviance explained in the present study appears adequate such that meaningful interpretations on the trends of the radial range data over time can be made.

In looking at the syllable pattern and difficulty level, an observed trend is noted for the difficulty level, but the overall pattern for the GAMs for vocabulary words does not seem to be entirely consistent compared to the results from the previous studies. One token per difficulty level is taken from the GAM plots to demonstrate the observed trend in the vocabulary word data, shown in Figure 5.4. The plots are arranged from top to bottom in order of increasing difficulty level.

Looking at the first few levels shown in Figure 5.4, there is a general trend where the radial distance data
seems to increase over time. As the vocabulary words become more difficult, however, it is observed that the radial distance data tends to decrease over time instead of increase. While this result does imply that difficult vocabulary words do impact balance, the balance is not impacted in a way that one might expect. In other words, while one might have expected the contrary based off of the results from the previous studies in this thesis, in this experiment it is observed that more difficult vocabulary words tend to lead to a decrease in slope of the radial distance of the COP.

The full set of plots for all the words are in Figure 5.5. In the figure, the row number represents the difficulty level (first row is level 1, second row is level 3... last row is level 21), and the column number represents the syllable structure (first column is CVCV, second column is CVCVC, third column is CVCVC (alliterated), fourth column is VCCVC, and fifth column is CCVCVC).

**Summary of the experiment on vocabulary words**

Overall, sway is affected by the content of the vocabulary words, where in particular there is an effect of difficulty level and a marginal effect of syllable type. Vocabulary words that are more difficult tend to have a larger fiftieth percentile of radial range, and although for the GAMs the trend isn’t exactly consistent with the other studies’ results so far, the more difficult vocabulary words tend to elicit more negative slopes of the radial distance data. These results from the fiftieth percentile of the radial range of the COP generally support the hypothesis that difficult language affects the postural sway, where language can effectively perturb standing balance, although more discussion may be required, particularly as to why the trends of the radial distance data differ compared to the other studies’ results. The next chapter is devoted to an overall discussion of the results, where this particular result of the GAMs and where it situates itself along with the other results from the other studies as a whole will be discussed in more detail.
Chapter 6

Discussion and interpretation of the findings from the four studies on balance and language

Overall, the main goal of the studies undertaken in this thesis was to investigate whether postural sway is affected by language tasks, whether in the form of paragraphs, sentences, or vocabulary words. To address this question, the findings from each of the four studies are discussed below, interpreting the results attained from each of the studies.

6.1 Paragraphs study overall review

The results of 2 suggest that there are differences in standing postural sway reflecting the difficulty level of paragraphs when subjects are listening to these paragraphs while standing. This general finding is true for both the fiftieth percentile of the radial range of the COP, as well as in the visual inspection of the trends from the GAMs observing the radial distance data over time.

6.1.1 Paragraphs and the fiftieth percentile of the radial range of the COP

The fiftieth percentile of the radial range of the COP is used to evaluate the overall trends subjects have when processing the tokens for the experiment on paragraphs. This parameter is used to get a general sense as to whether balance changes depending on the difficulty level, and if it changes in a way that one would believe to be related to changes in cognitive load.

In the first study, the study on paragraphs shows a difference with regards to difficult passages having larger range of overall sway compared the easier and the medium passages. This finding can be interpreted as understanding that postural sway is sensitive to passage difficulty, and this finding makes sense in the way that one might expect, in that people would sway more and not less when faced with challenging passages. Difficult passages elicit greater changes in postural sway, which would imply that postural sway reflects changes in cognitive load during language processing of paragraphs while standing.

The medium and easy passages don’t statistically differ from each other in the post-hoc tests, which perhaps implies that native speakers don’t process the two types of passages a way that would reflect in
their postural sway. Subjects’ sway patterns to difficult passages may be more sensitive to the material than easier passages. Alternatively, the difficult passages were designed to be challenging for college aged adults, but the easy and medium passages may be perceived to be the same overall difficulty. Cognitive resources required to comprehend such passages would not reflect in their balance. A college student may have an easy time at anything below their level of competency such as grade school and high school level writing, which would be reflected by the lack of any relative difference in their postural sway patterns.

On the other hand, however, while the similarity in the results from the easy and medium passages may imply that a threshold may exist for which only certain difficult passages can elicit changes in postural sway, the results from this study might be also linked with the fact that the material for the difficult passages are created differently than the easy and medium passages. In other words, using an algorithm like SCIgen for the easy material and medium material might make the differences in postural sway more apparent for the easy and medium passages. If the passage materials were created using an algorithm similar to SCIgen but instead used simpler words and grammatical structures, the corresponding paragraph materials, while still easier than the difficult passages, would be much more unfamiliar and new to the subject. Instead of using writing samples from grade-school and high school texts, subjects would have been exposed to unfamiliar material, while still having simpler words and grammatical structures than the difficult passages. This design would be an interesting path to follow in a future study.

6.1.2 Paragraphs and the radial distance of postural sway using GAMs

To get a visual sense of the subjects’ sway patterns as they listen to each of the stimuli, the tokens’ radial distance data as a function of time is observed. A fitted set of splines for each token respectively for the radial distance of the COP is generated by using the GAMs. The resulting plots from the GAMs provide a way to better understand the trends or patterns of the sway data through visual inspection.

The different slopes that are generally present for each of the respective difficulty levels can make sense if the radial distance data is thought of as reflecting changes in cognitive load. For example, for passages of easy difficulty, the radial distance is observed to generally slope downwards throughout the duration of each token. Importantly, the radial distance is affected in a different manner over the duration of the tokens when the passages become more difficult, such that the overall slope is increasingly greater when the medium or difficult passages are encountered, respectively. In other words, while the subjects have a general downwards trend of sway over time when processing paragraphs, as the passages become more difficult, the sway is more resistant to this downwards trend.

Therefore, if one thinks of the radial distance of postural sway as reflecting changes in cognitive load,
this difference in slope would imply that changes in cognitive load impact postural sway, where the overall interpretation is again consistent with the general theme of the thesis. Postural sway is affected by the language difficulty, and the more difficult passages elicit changes in cognitive load such that during the passage, subjects will move in a way that is intuitively reflective of greater amounts of thinking. The results suggest that when subjects listen to, for example, an easy passage, subjects will have a certain pattern of sway, but this pattern changes when the the material becomes more difficult. As the passages become more difficult, the relative slope of the radial distance increases, which may imply that more thinking is involved when compared to the easy passages, as it manifests in the differences in their standing balance.

Overall, in thinking about both the results from the fiftieth percentile of the radial range of the COP and how the radial distance data varies over time across the paragraph tokens, the results suggest that postural sway does seem to be affected by paragraph difficulty. The next section will discuss the results from the HNPS and SS/SO sentences.

6.2 HNPS and SS/SO study overall review

In this study on sentences, HNPS sentences and SS/SO relative clauses were used in order to test the effect of sentence processing on postural sway. These two sentence types were chosen because they both satisfied two requirements, the first being that the sentence type must have some sort of easy and difficult variant, and the second being that the difference between the easy and difficult variant must depend on only changing the word order. This second requirement is important in the sense that by testing sentence pairs that only differ in their word order, significant findings in the results will be more likely due to grammatical differences in the sentences and not vocabulary word differences.

The results from the study on sentences suggest that subjects sway more when the sentence is relatively more challenging to process. Similar to the results on the paragraphs, postural sway is affected by the processing of sentences, and this finding is demonstrated in both of the tested sentence types of HNPS and SS/SO sentences, which is discussed in the following sections.

6.2.1 HNPS and SS/SO sentences and the fiftieth percentile of the radial range of the COP

The fiftieth percentile of the radial range of the COP is used to evaluate the overall trends subjects have when processing the tokens for the experiment on sentences. This parameter is used to get a general sense as to whether balance changes depending on the difficulty level, and additionally, if it changes in a way that
one would believe to be related to changes in cognitive load.

For this study on sentences and its potential to affect postural sway, both of the sentence types exhibit traits in balance that are consistent with the idea that subjects sway more when the sentence type is relatively more difficult. For the HNPS sentence types, the acceptable version of the sentence incurs less amounts of postural sway relative to its less acceptable counterpart. Similarly, for the SS/SO sentence types, the SS sentence type incurs less amounts postural sway than the SO sentence type, as reflected by the fiftieth percentile of the radial range of the COP.

These results make sense if one thinks of the postural sway as reflecting changes in cognitive load. The implication would be that postural sway correlates with changes in cognitive load, such that when subjects listen to a given sentence, their overall range of sway is increased when the sentence should be relatively more difficult to process. Therefore these results are also promising in the sense that like the study on paragraphs, there is a clear difference between easier and harder sentences which is manifested in the form of changes in balance.

6.2.2 HNPS and SS/SO sentences and the radial distance of postural sway using GAMs

To get a visual sense of the subjects’ sway patterns as they listen to each of the stimuli, the tokens’ radial distance data as a function of time is observed. A fitted set of splines for each token respectively for the radial distance of the COP is generated by using the GAMs. The resulting plots from the GAMs provide a way to better understand the trends or patterns of the sway data through visual inspection.

Again, the data makes sense if one thinks of the radial distance of postural sway as reflecting changes in cognitive load. If one believes that postural sway is linked with changes in cognitive load, the majority of the sentence tokens would then support the same general concept that an increased slope leads to an increased amount of thinking compared to its relatively easier counterpart. For the HNPS tokens, in the regions where the tokens have differed in their word order, higher slopes are present for the majority of tokens in GAMs relative to their more acceptable counterparts. Likewise, for the majority of the SS/SO tokens, higher slopes are present for the SO tokens in the regions where the difference between the word order is present, which should be the case as SO sentences are reported in the literature to incur larger amounts of cognitive load than SS sentences. Overall, the implication is that for sentences, the regions where more thinking is involved reveals itself in the form of increased postural sway over time as represented by the radial distance data.

Therefore, in thinking about both the results from the fiftieth percentile of the radial range of the COP and the how the radial distance data varies over time across the trials, postural sway does seem to be affected
by sentence difficulty, which indicates that postural sway can be used to assess changes in cognitive load for sentences. So far, postural sway shows promise in both paragraphs and sentences for its ability to reflect changes in cognitive load, and the next section will discuss the results from the verification study.

6.3 Verification study overall review

The postural sway analysis on the tokens used from [Juzek, 2016] provides some insight for how studies on postural sway and language acceptability may be related. In this verification experiment, there is a strong correlation between postural sway data and sentence acceptability judgments when syllable count is normalized when utilizing postural sway. In addition to this finding, similar to the overall results from the paragraphs and sentences in the previous studies, there does seem to be an implication of a language effect on postural sway with the tokens tested.

6.3.1 Verification study sentences and the fiftieth percentile of the radial range of the COP

The fiftieth percentile of the radial range of the COP is used to evaluate the overall trends subjects have when processing the tokens for the experiment. This parameter is used to get a general sense as to whether balance changes depending on the difficulty level, and if it changes in a way that one would believe to be related to changes in cognitive load.

Postural sway is potentially affected by sentence type when syllable count is accounted for, which suggests that there is a link between language and changes in cognitive load for the token set used by [Juzek, 2016]. In other words, if one factors in syllable count, the results suggest that postural sway may be affected by the relative difficulty levels of the tokens. The result once again bears similarity to the previous studies in understanding the relationship between postural sway and language difficulty. In thinking about how changes in cognitive load may affect balance, subjects may end up swaying more when the sentence token is more difficult.

6.3.2 Verification study tokens and the radial distance of postural sway using GAMs

To get a visual sense of the subjects’ sway patterns as they listen to each of the stimuli, the tokens’ radial distance data as a function of time is observed. A fitted set of splines for each token respectively for the
radial distance of the COP is generated by using the GAMs. The resulting plots from the GAMs provide a way to better understand the trends or patterns of the sway data through visual inspection.

The data from this experiment makes sense if one thinks of the radial distance of postural sway as reflecting changes in cognitive load. The results from the GAMs help to illustrate the possibility that in places where difficulty might be expected, the slope of the radial distance generally increases. In the regions where one might expect postural sway to be affected, the radial distance generally deviates in those particular regions, such that in thinking about the possible link between language and changes in postural sway, regions in time where the radial distance data rises in slope generally corresponds to the same regions where the sentences should theoretically require more cognitive load to process. Therefore, the results from the GAMs suggest that postural sway can be thought of as being representative of changes in cognitive load. The data makes sense if one believes that the trends from the GAMs suggest that increased slopes of the radial distance data are indicative of larger amounts of thinking involved. Overall, the GAMs show that the variation of the radial distance data over time seem to correspond to changes in cognitive load, where the regions that have incurred more changes in cognitive load are reflected in the fact that subjects have a relatively greater amount of sway.

In looking at both the GAM results from this verification study and the study on HNPS and SS/SO sentences, both results support the possibility that postural sway can be used to assess changes in cognitive load at specific points in time. The trends generally correspond to the regions in time where the relative increase in cognitive load is expected, such that changes in the radial distance data occur approximately around the place where the difficulty is expected. Prior to the experiments, we had theorized that the effect of language difficulty would reveal itself in changes in postural sway over time, and particularly that this trend would show up either in sync or with some amount of time lag to the actual regions where the difficulty in the audio would have occurred. Because both SS/SO and HNPS sentences and the sentences used in the verification study show trends in the data that at least generally match the regions of theoretical changes in cognitive load along the time axis, it may be that for sentences in general, that meaningful information related to specific regions in time can be extracted from the trends of the postural sway.

In other words, if an increase in radial distance is generally in the location where the difficulty in the sentence should be occurring, then when the subject encounters a difficult part of the sentence, the sway is reacting in a way that may be meaningful in that given region of time. This region where the change in sway occurs can help provide an understanding when a person has difficulty understanding a sentence, which is a useful finding because we theorized that this particular trend would have shown up in the results. To more objectively determine whether there is a lag between sway and thinking for sentences, future studies
can be done to determine the degree to which the radial distance data matches the time from which actual electrical signals come from the brain (for example, in comparing results with pre-established tools that measure brainwave electrical activity such as EEG) when processing sentences.

Overall, in thinking about both the results from the fiftieth percentile of the radial range of the COP and how the radial distance data varies over time across the trials, postural sway does seem to be affected by sentence difficulty from the tokens used in [Juzek, 2016], which further establishes the connection between postural sway and changes in cognitive load. So far, all the findings suggest that in addition to paragraphs, postural sway also shows some promise in evaluating sentences so long as syllable count is accounted for. The next section will discuss the results from the vocabulary study.

6.4 Vocabulary study overall review

The study on vocabulary words have somewhat differing results when compared to the previous other studies on sentences and paragraphs, but some useful information is provided regarding the extent for which postural sway can be used to quantify changes in cognitive load. There are also reasons suggesting that words may be too short and that it has reached the limit of the technology, and/or the repetition used in the experiment to make the stimuli longer has introduced artifacts in postural sway. Results suggest that word frequency and possibly syllable structure can affect postural sway. While it is noted that trends do occur for the radial distance data over time under the GAMs, the visual results are not entirely consistent with the rest of the studies, suggesting perhaps some kind of limitation with regards to utilizing postural sway as an intuitive means to measure changes in cognitive load due to language processing. These results are discussed in the following sections.

6.4.1 Vocabulary study and the fiftieth percentile of the radial range of the COP

The fiftieth percentile of the radial range of the COP is used to evaluate the overall trends subjects have when processing the tokens for the experiment. This parameter is used to get a general sense as to whether balance changes depending on the difficulty level, and if it changes in a way that one would intuitively understand to be related to changes in cognitive load.

In thinking along the lines that changes in postural sway correspond to changes in cognitive load, unlike the other three studies on postural sway and language processing, the results are inconsistent with previous experiments. However, the post-hoc tests do show differences in the difficulty levels. The overall results from
the fiftieth percentile of the radial range of the COP suggests that higher difficulty levels have higher amounts of sway, and that there may be some differences present within the syllable structure. The implication would be then that changes in cognitive load are reflected in the subject’s balance for vocabulary words, but again, this effect is not as strong or clear as the other studies involving complete sentences and full paragraphs.

The lack of significance between syllable structures, and a weak difference between the difficulty levels may make sense in understanding the overall context of the present thesis, where between the three topics of paragraphs, sentences, and vocabulary words, that vocabulary words are the smallest element and therefore the most challenging out of the three. Testing just vocabulary words, perhaps, may provide the most challenge in terms of extracting a clear and useful relationship in understanding postural sway and changes in cognitive load due to language processing.

Overall, for the fiftieth percentile of the radial range of the COP, the analysis on the vocabulary words suggests that there is a relationship that exists between language and balance, in that changes in cognitive load is related to corresponding changes in postural sway. It is noted, however, that this finding is not necessarily as strong per se as the studies done in this thesis on sentences and paragraphs. Nevertheless there does exist a similar trend in the vocabulary words compared to the other studies, which does merit the possibility of utilizing the fiftieth percentile of the radial range of the COP to carry out the analysis of vocabulary words.

6.4.2 Vocabulary study tokens and the radial distance of postural sway using GAMs

To get a visual sense of the subjects’ sway patterns as they listen to each of the stimuli, the tokens’ radial distance data as a function of time is observed. A fitted set of splines for each token respectively for the radial distance of the COP is generated by using the GAMs. The resulting plots from the GAMs provide a way to better understand the trends or patterns of the sway data through visual inspection.

In observing the trends of the radial distance data over time, once again in this study on vocabulary words, in thinking along the lines that changes in cognitive load correlate with changes in postural sway, the results are more challenging to interpret compared to the other three studies. Although the observed pattern thus far has been that greater difficulty amounts to larger slopes in the radial distance data, the trend from the results on the vocabulary study is somewhat reversed. The more easy vocabulary words elicit relatively higher overall slopes than the difficult vocabulary words, which goes against the trends of the previous studies. Therefore the results from the GAMs suggest that larger sway does not necessarily reflect larger changes in cognitive load when focused on the elements of language smaller than sentences.
In other words, there is a general trend where the easy vocabulary words have a higher slope compared to the more difficult vocabulary words, but this trend is the opposite of what is being shown in the studies on sentences and paragraphs. It may be the case that the size of the tested elements can be problematic, in that vocabulary words may be too small for postural sway to clearly represent changes in cognitive load. Therefore, one possible interpretation of the results is that there is not a straightforward relationship between the magnitude of the effect of language difficulty on sway when it comes to smaller elements such as vocabulary words, where easy or difficult vocabulary words provide too small an effect for postural sway to be affected in the same way that the standing sway from sentences and paragraphs are.

However, one factor to consider is to recall that in the vocabulary experiment, the words were repeated a total of six times for each token in order to extend the length of time for which postural sway can be collected. This repetition was done over a concern that collecting data on single words would not offer an adequate amount of postural sway to analyze, given the fact that not only are single vocabulary words being analyzed, but that postural sway literature suggests that at least 60 seconds are required for reliable results ([Doyle et al., 2007]). This repetition difference between the study on vocabulary words and the rest of the studies on sentences and paragraphs may also contribute to the fact that the trends in the GAMs are not consistent with the findings from the rest of the studies.

To sum, the results from the study on vocabulary words suggest that postural sway can be used to measure changes in cognitive load when using the fiftieth percentile of the radial range, but that while a trend does exist for the GAMs, it’s harder to claim that the same relationship between changes in cognitive load and language difficulty exists when compared to the results from the studies on sentences and paragraphs. The results from the GAMs are somewhat more challenging to interpret relative to the other studies in regards to establishing a firm link between standing balance and changes in cognitive load, which may imply that a threshold or limit is reached when testing just vocabulary words on standing balance. Compared to the other three studies, the results from the GAMs follow something of an opposing trend when compared to the studies on sentences and paragraphs for the radial distance data over time.

All in all, in thinking about the link between postural sway and changes in cognitive load for the processing of vocabulary words, a link does exist, but when thinking about the trends of the GAMs, this relationship is less intuitive to interpret, for which the relationship found through the GAMs may still be meaningful, but the results are not possibly as fruitful from the perspective of identifying a useful trend in language processing and balance, where larger amounts of sway would equate to larger amounts of changes in cognitive load.

However, between the four studies, there is a valuable lesson learned in understanding not only that postural sway can be used to measure changes in cognitive load during language processing, but also, how
far one can push this analogy in terms of intuitively understanding the results of the sway data. In other words, in completing this study on vocabulary words, a general sense as to how far postural sway can intuitively reflect changes in cognitive load has been attained, which is a useful finding in and of itself for any future studies that may deal with postural sway and relative difficulty levels involved in language processing, the overall impact of which is discussed in more detail in the following section.

6.5 Overall impact from the four studies

6.5.1 Testing paragraphs, sentences, and vocabulary words

Overall, the results suggest that postural sway is sensitive to the processing of paragraphs, sentences, and possibly vocabulary words. Therefore, postural sway may be able to be used for measuring changes in cognitive load during language processing, where generally speaking, larger amounts of thinking relates to larger amounts of sway. The results also suggest that some care needs to be taken with regards to the size of the element being measured, particularly for single words, where some difficulty is encountered in interpreting the results in the study on vocabulary words, particularly from the results of the GAMs. For paragraphs and sentences, a more clear relationship is found, but for vocabulary words, a conflicting trend in the data seems to have been reached.

6.5.2 Vocabulary words as the limiting factor

If one desires to use postural sway to measure changes in cognitive load during language processing, the fact that vocabulary words by themselves seem to present a limitation on the intuitive interpretation of postural sway during language processing does make some sense in that among the three tested, the smallest element, vocabulary words, is the one that is hardest to interpret. Therefore, the findings from the study on vocabulary words are useful to know in implying that for studies focusing on postural sway and language, that future studies should probably focus on elements larger than vocabulary words. It may be the case that vocabulary words by themselves do not elicit enough of a change in postural sway such that an intuitive trend can be found. If one were to make a broader conclusion based off the results of all four studies, then it would seem that vocabulary words are the element for which postural sway can least reliably measure changes in cognitive load.

It is noted that the repetitions were done because of a concern that not enough postural sway data would be collected for a robust analysis, not only because single vocabulary words by themselves don’t take a lot of time to listen to, but that 60 seconds are recommended to have reliable postural sway results
However, in light of this design difference between the vocabulary word study and the sentence and paragraph studies, the repetitions may be altering the results of the vocabulary study in relation to the other studies on sentences and paragraphs.

Therefore, to address the question posed in the beginning of the thesis, the results suggest that it is fairly possible for postural sway to measure changes in cognitive load for paragraphs and sentences. Differences in postural sway are observed in the experiment on vocabulary words, but the interpretation of the results turns out to be somewhat different than the other studies, particularly from the results on the GAMs, and this is an important limitation to note. Keeping this limitation in mind, the overall lesson that has been obtained in carrying out all four studies is that postural sway can be utilized to measure changes in cognitive load during language processing.

### 6.6 Alternative methods of measuring changes in cognitive load during language processing

Taking all the studies together under consideration, the results suggest that postural sway may provide a way to measure changes in cognitive load. For paragraphs and sentences, the studies’ results suggest a link between postural sway and changes in cognitive load, such that more difficult items relate to more changes in postural sway, whether it’s the fiftieth percentile of the radial range, or how the radial distance data varies over time for each token. The final results suggest that this general trend is present for at least the elements of language that are greater than vocabulary words.

In closing, the present thesis successfully finds one way to measure changes in cognitive load during language processing, which may be useful in language studies that would involves utilizing such measures of cognitive load in an efficient manner by measuring the characteristics and movements as to how people stand. The next section will talk more about postural sway and how balance research may situate itself relative to other methods of evaluating changes in cognitive load during language processing. Alternative methods of measuring changes in cognitive load along with their pros and cons are discussed below.

#### 6.6.1 Understanding where measuring balance stands relative to other methods

The experiments in this thesis have established a relationship between postural sway and difficulty related to language processing using force plates, and it’s reasonable to infer that other measurements from different technologies will show this relationship as well. One important topic is noted in that measuring balance
would just be one of several ways to measure changes in cognitive load. There are other alternatives available, and it is worthwhile to take a moment to evaluate where balance measures can stand relative to the other technologies that researchers have at their disposal.

Other methods besides measuring balance that can measure changes in cognitive load include methods such as tracking the rapid movements of the eye using eyetrackers, measuring electrical activity on the scalp that comes from the brain through EEG, or interpreting images of blood flow in the brain with fMRI, which are discussed in more detail in the following sections.

Advantages of the other methods

Electroencephalography (EEG) is a method that involves tracking and recording electrical activity from the brain. By placing multiple sensors on the subject’s scalp, the electrical activity exiting the scalp can be recorded, where interpretation of the data can be made for example, to determine when a particular section of a sentence invoked greater amounts of cognitive load, and EEG has been utilized often in the analysis of sentence processing (e.g., [Kim and Osterhout, 2005], [Kolk et al., 2003], [Van Herten et al., 2005]).

One primary advantage EEG has over postural sway data is in the physical position for which this data is collected. In other words, the electrical activity recorded from the sensors on the scalp is very close to the brain itself. Relative to balance recording equipment such as from the force plate, the actual distance between the source (the brain) and the recording instrument (the force plate) is something that EEG has a definitive advantage in. EEG also samples the data at high sample rates, such that the rapid processing of language can be sufficiently captured.

Functional magnetic resonance imaging (fMRI) is arguably even closer than EEG in detecting localized regions of brain activity. Assuming that higher brain activity relates to higher metabolism, if an area in the brain is relatively more active, then it requires more oxygen, and oxygen is supplied to the brain via increased blood flow. fMRI records the patterns of blood flow in the brain, and consecutive snapshots of the head can be taken to observe cognitive activity during language processing, and the analysis of word and sentence comprehension is also commonly done in language fields using fMRI ([Price, 2010]).

One primary advantage of fMRI is also in its ability to very closely monitor the brain, closer than what measuring balance can. fMRI can inform the researcher what regions of the brain have been activated when processing sentences, which is something that measuring postural sway cannot easily achieve.

Eye trackers are tools that track the motion of the eye. An infrared light is sent to the pupil, and the resulting reflection is tracked with computer software, allowing for an understanding of how the eye moves when observing a picture or reading a sentence. Eye movements are categorized as either fixations and sac-
cades (rapid eye movements between fixation points). Language research utilizes eye trackers to understand language processing during, for example, sentence comprehension, and observes the saccades to get an understanding of the subject’s thought processes through analysis of their eye movements ([Rayner et al., 2012]).

One advantage of eye trackers is that the movements of the eye allow researchers to better understand what the subject is thinking. Eye trackers have a sufficient sample rate to capture the data of the eye movements. The specific advantage it has over postural sway is that it allows the researcher to more accurately assess what the subject was thinking at a particular point in time. For example, in a complicated text passage, the eye tracker will provide information as to how long a given sentence or set of words was fixated on, and if he or she backtracked to that region after giving the passage a quick read. Data from postural sway can only provide the general patterns of changes in cognitive load, but not the specific parts of the passage that he or she may have been thinking about.

Having covered these three alternative methods to assessing changes in cognitive load, though, there are certain advantages that collecting postural sway data has as well. These points are discussed in more detail below.

**Advantages of measuring balance**

One advantage that utilizing postural sway has over some of the other technologies is that measuring balance provides an acquisition of its necessary data relatively quickly. For instance, setting up a study that uses EEG takes considerably more preparatory time than preparing, for example, data collection on a force plate. Subjects who participate in an EEG study have to have their head washed and cleaned in order to elicit the best physical connection between the sensors and the scalp. The subject’s head is washed so that a gel can be placed on their head for the sensors to be placed on, and it not uncommon to have approximately 30 minutes or more of preparatory work prior to data collection. Using force plates, in contrast, does not require any considerable amount of prep work besides turning on the equipment.

Furthermore, another advantage is that recording postural sway is relatively inexpensive compared to some of the other technologies. Compared to eye trackers, EEG, and fMRI, measuring postural may provide the most economical solution towards attaining equipment that can quantitatively measure changes in cognitive load during language processing. Devices that measure postural sway can be had for as low as 15 dollars, and there are other ways to collect postural sway data besides using force plates, such as using built-in sensors that come with modern day cell phones. The potentially low cost involved in getting devices that can record postural sway would be useful in the way that if desired, thousands of dollars or more worth of equipment is not a necessity, if the researcher is interested in investigating the most economical path for
him or her to take regarding the attainment of data dealing with changes in cognitive load during language processing.

Additionally, measuring balance has an advantage over fMRI, in terms of physical positioning and frame rate. Subjects in fMRI research are laying in the supine position. Because they are lying down, depending on the situation desired, it may be more difficult to capture data in a natural setting using fMRI, considering that the region where the subject is situated in is relatively more space restricted than they would be with tools measuring postural sway. Furthermore, fMRI sampling rates are generally on the order of once every few seconds ([Ludden, 2015]), for which measuring balance again has the advantage of with regards to temporal resolution.

Overall, the advantages of measuring balance may provide a way to meaningfully position itself in the research fields as a useful alternative to measuring changes in cognitive load during language processing. Utilizing postural sway to measure changes in cognitive load during language processing can be yet another tool in the arsenal of the researcher, for which if chosen as the tool for the right job, the demands of the language research can be more conveniently and efficiently met. While not a direct measurement of cognitive load, measuring postural sway may be accurate enough for many linguistic purposes.
Chapter 7

Conclusions

In this thesis, four experiments are done to examine the effect of language difficulty on postural sway. One experiment was done on paragraphs, two experiments were done on single sentences, and one experiment was done on vocabulary words. The overall premise is that in using the dual-task paradigm, because cognition and balance utilize a shared pool of cognitive resources, that changes in balance reflect changes in cognitive load. Subjects are asked to listen to either complete passages, sentences, or vocabulary words while standing.

The main chapters of the present work provide evidence supportive of a possible relationship between postural sway and language processing, and for each of the studies, the trends present themselves in a relatively consistent manner. To summarize the first study, a relationship between sway and difficulty level for paragraphs are present, particularly between the easier and hardest levels. Likewise, for sentences in the second study, there is an effect of difficulty level on sway for HNPS and SS/SO sentences. For the work on verification in the third study, there is a strong correlation in sway and acceptability if syllable count is taken into account. When observing the radial distance of the COP as a function of time in each of these studies, GAMs show that generally speaking, more difficult tokens are higher in slope than their non-difficult counterparts, which further suggests a link between standing balance and cognitive load. Lastly, for vocabulary words in the fourth study, there is an effect of difficulty on sway for some of the levels, but in the analysis of the radial distance data over time, the trend is not the same as the other studies on sentences and paragraphs, where an increasing slope with increasing difficulty level is not observed. For vocabulary words, a limit may have been reached with regards to intuitively understanding and utilizing the trend of postural sway as it relates to language processing.

Overall, however, this thesis demonstrates that when individuals are subjected to difficult language, particularly sentences and paragraphs, that they may sway more. This conclusion is attained by using the dual-task paradigm, where because cognition and balance utilize a shared pool of cognitive resources, changes in balance reflect changes in cognitive load, giving a way to meaningfully interpret and understand standing balance from a linguistic standpoint.

Going forward with possible paths for future research, an effort can be placed on streamlining the data
collection and analysis for the development of practical applications. Having focused on establishing a link between language processing and changes in cognitive load, a long-term study worth considering is to determine whether there is a way to attain balance measures in advance for any type of sentence. In other words, rather than have an individual research study performed for every given sentence type, is there a way to predict the difficulty level of text based on a corpus of postural sway data, and if so, how to create a large database of postural sway for this purpose?

It is our hope that a successive completion of a set of studies would pave the path for a relatively fast and useful inference to changes in cognitive load for language processing, such that a numerical quantity based on postural sway can be provided for any arbitrary sentence. An individual would query about a sentence into an online webpage, where a numerical result based off of large amounts of pre-collected postural sway data could then be returned. If the studies in the present thesis are to be extended for the sake of creating a larger database of postural sway as it relates to language, then the following venues for further investigation are recommended, in the following order:

1) Search for a way to efficiently obtain postural sway data for language processing by taking advantage of alternative technologies besides force plates. Is there is a reliable method for attaining postural sway aside from the use of force plates as done in the present research? For example, most modern-day smart phones come with accelerometers built into themselves, and accelerometers have the potential to track postural sway if properly fixed on the subject’s body. Although the type and build of accelerometer would vary from phone to phone, accelerometers are an appealing example of an alternative technology when compared to force plates in their being more commonplace, small, and transportable.

2) Investigate alternative methods to attain larger quantities of postural sway data in shorter amounts of time. For instance, is there a way to attain postural sway data outside of the research laboratory? Postural sway data is attainable in a lab, but more data could be acquired relatively quickly if it could be collected at, say, the comfort of the subject’s own home. For example, if the data of postural sway for a given sentence stimuli could be uploaded to an internet server, it would allow the collection of postural sway data for a large audience without the need of everyone traveling to a laboratory for testing.

3) Upon completion of the above two studies, attain postural sway data with a focus on a very large amount of different sentence types. In other words, test a large number of various types of grammatical and ungrammatical sentences, and record the postural sway, uploading it into one large online and publicly accessible repository.

4) Investigate methods for getting results on postural sway on untested sentences with machine learning algorithms. If a large enough database is created for various grammatical and ungrammatical sentences, one
could attempt to gain an understanding of general trends of postural sway for sentences not captured in
the database by recombining relevant components of a given sentence. Such a feat could be accomplished,
for example, by using the database to approximate how much each relevant component contributes to the
overall difficulty of each sentence.

The present research focuses on standing balance as it relates to changes in cognitive load during language
processing. The hope is that these results presented in this thesis will serve as a starting point for setting a
larger research framework, for which one can use to efficiently quantify changes in cognitive load associated
with language processing by utilizing changes in postural sway.
References


