The effects of pausing on comprehensibility

Rebecca M. Bae
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd
Part of the Linguistics Commons

Recommended Citation
Bae, Rebecca M., "The effects of pausing on comprehensibility" (2015). Graduate Theses and Dissertations. 14324.
https://lib.dr.iastate.edu/etd/14324

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
The effects of pausing on comprehensibility

by

Rebecca Bae

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

Major: Teaching English as a Second Language/Applied Linguistics (Language Assessment)

Program of Study Committee:
John Levis, Major Professor
Gary Phye
Volker Hegelheimer

Iowa State University
Ames, Iowa
2015

Copyright © Rebecca Bae, 2015. All rights reserved.
DEDICATION

For my family
# TABLE OF CONTENTS

DEDICATION .................................................................................................................. ii

LIST OF FIGURES ........................................................................................................... v

LIST OF TABLES .............................................................................................................. vi

ABSTRACT ...................................................................................................................... vii

CHAPTER 1. INTRODUCTION ......................................................................................... 1

CHAPTER 2. LITERATURE REVIEW .............................................................................. 4
  2.1 Measures of Listeners’ Perceptions ................................................................. 4
  2.2 Pausing ................................................................................................................. 7
  2.3 Perceptions of Pausing ...................................................................................... 8
  2.4 The Current Study ............................................................................................. 10

CHAPTER 3. METHODOLOGY ..................................................................................... 12
  3.1 Materials ............................................................................................................ 12
     3.1.1 Participants ................................................................................................. 12
     3.1.2 Speech Sample ........................................................................................... 14
     3.1.3 Recordings .................................................................................................. 15
  3.2 Procedures ......................................................................................................... 23
  3.3 Analysis .............................................................................................................. 24

CHAPTER 4. RESULTS .................................................................................................. 27
  4.1 Effects of Pause Treatment on Comprehensibility ......................................... 27
  4.2 Comprehensibility vs. Fluency ........................................................................ 30
  4.3 Kang’s Modified Instrument .......................................................................... 31

CHAPTER 5. DISCUSSION ............................................................................................ 33
  5.1 Additional Comprehensibility Items ............................................................... 39
     5.1.1 Comprehensible/Incomprehensible ......................................................... 39
     5.1.2 Clear/Unclear ............................................................................................ 41
     5.1.3 Required little/lots of effort and difficult/simple to grasp ...................... 41

CHAPTER 6. CONCLUSION ........................................................................................... 43

APPENDIX A. TRANSCRIPT .......................................................................................... 47

APPENDIX B. PAUSE ANALYSIS BETWEEN NATIVE AND NON-NATIVE PAUSES .... 49

APPENDIX C. VERSION A – NORMAL RECORDING WITH NATIVE-LIKE PAUSES ......... 52
LIST OF FIGURES

Figure 1. Comprehensibility instrument from Kang (2010) ......................................................... 6

Figure 2. Two possible Latin squares ................................................................. 22
LIST OF TABLES

Table 1. *Classification of pausing according to length* .......................................................... 7

Table 2. *Demographics of the participants and the university* ........................................... 13

Table 3. *Distribution of participants across classes* ................................................................. 13

Table 4. *Distribution of pauses across the sections of the lecture* ...................................... 18

Table 5. *Class recordings with randomized distribution of the four sample recordings* ..... 23

Table 6. *Effect tests of the three-way ANOVA* ................................................................. 27

Table 7. *Means and standard deviations for the different treatments by class* ................. 28

Table 8. *Post-hoc Tukey test for pause treatment* .............................................................. 29

Table 9. *Spearman correlations for comprehensibility and fluency* ................................. 30

Table 10. *Spearman correlations between “easy/hard to understand” and the remaining items* ................................................................. 32
ABSTRACT

Pausing effects how well listeners understand and attend to meaning in discourse. This study investigates the effects of three different pause treatments (irregular placement, increased frequency, and longer length) on comprehensibility ratings. Varonis and Gass (1982) found that a complex interaction of factors affected comprehensibility ratings. These included pronunciation, grammar, familiarity and fluency. While many of these features have been investigated (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997; Hahn, 2004; Isaacs & Trofimovich, 2012; Kang, 2010; Munro & Derwing, 1995), pausing has received little attention. In this study, an extended NNS speech sample with native-like pausing was manipulated, creating three experimental recordings, one with irregularly placed pauses, one with increased pause frequency, and one with longer pauses. Forty-three undergraduates in four different class groups listened to each of the pause treatments and rated them for comprehensibility. In addition to comprehensibility measures, participants also rated each treatment for fluency based on Griffiths’ (1991) proposition that pausing is often tied to fluency. This allowed for a comparison of the effects of pausing on comprehensibility to those on fluency. Additionally, this study investigated the strength of Kang’s (2010) revised comprehensibility instrument. The results showed that irregular pause placement was the greatest detriment to comprehensibility, followed by pause frequency. These results may be explained by the psycholinguistic model of language processing which assumes we process language in chunks. When NNSs pause irregularly, NS listeners must process each word individually to make meaning instead of processing the chunk through expected phrasing. This, in turn, causes lower comprehensibility ratings. These results advocate for the teaching of formulaic language in the ESL/EFL classroom so that attention
to pause placement in conjunction with work on fluency facilitates more comprehensible speech.
CHAPTER 1. INTRODUCTION

Speaking is a two-way process. The speaker strives to make a clear, succinct utterance, and the listener must focus on processing the information and creating understanding. While the produced speech is clear evidence of the speaker-oriented process, there is no equivalent for the listener-oriented process. There are no outward manifestations of the listening itself; however, there are variables that can be investigated, such as comprehensibility. Comprehensibility is the listener’s perception of how easy it is to understand an utterance or piece of spoken discourse (Isaacs & Trofimovich, 2012).

Understanding comprehensibility is particularly important for the TESL community. If we can ascertain what makes someone easy or difficult to understand, we can make more informed decisions as teachers and assessors. Stated another way, if we can determine whether grammaticality, stress, pitch, segmental production, fluency, etc., cause disruption in understanding on the part of the listener and to what extent, then we can focus on the most salient features of speech in the EFL/ESL classroom (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997; Kang, 2010; Munro & Derwing, 1995; Varonis & Gass, 1982). Comprehensibility ratings can also be used in other ways. For example, Derwing, Munro and Wiebe (1998) used comprehensibility ratings to determine which of three different pronunciation pedagogies was most effective over a 12-week course. They found that comprehensibility scores increased for students receiving global and prosodic instruction as compared to students receiving segmental instruction and no specific instruction. In another example, Isaacs and Trofimovich (2012) used comprehensibility ratings to create an oral assessment instrument. Vocabulary usage, word stress, mean length of run, story breadth, and grammatical accuracy were found to distinguish between different
comprehensibility levels and were thus used to create the instrument. Underlying all of this work is the idea that increasing a non-native speakers’ comprehensibility, their ability to be understood easily, is key to effective communication.

To investigate comprehensibility many studies look at multiple variables to ascertain their relative importance. Hahn (2004), however, looked at comprehensibility by isolating just one variable, that of primary phrase stress. Hahn created three different recordings, one with correct primary stress placement, one with misplaced primary stress, and one in which primary stress was absent. She had 90 university freshmen rate the three recordings (30 students for each recording) and found that the recording with correct primary stress was significantly more comprehensible than the other two recordings. The results also suggested that misplaced stress hindered understanding more than the lack of stress. This methodology allowed Hahn to focus on just one feature of speech production, and as such, she could investigate which aspects of primary stress affected comprehensibility. Similarly, this study will examine the effects of one variable on comprehensibility; however rather than primary stress, this study will focus on pausing.

Pausing is of particular interest because it is a salient feature of speech that has been investigated in previous research (Isaacs & Trofimovich, 2012; Kang 2010), and it is an area of study that both Griffiths (1991) and Riggenbach (1991) put forth as needing further investigation. However, despite Griffiths’ and Riggenbach’s call for more research in this area, there has been little research focusing solely on pausing in the intervening twenty-five years. Certainly pausing has been included as one feature of many, but it has rarely been the focus of dedicated study. To this author’s knowledge, Tavakoli (2011) is the only exception
to this general trend, and this one study alone cannot respond to the need for more
information about the effects of pausing on comprehensibility. This study fills this need.

Following the methodology of Hahn (2004), one controlled speech sample produced
by a non-native speaker was modified for pause placement, pause frequency and pause length.
This created four different treatments, one normal and three experimental, while controlling
for content, speaker, and accent. Participant groups listened to each of the different
treatments and rated them for comprehensibility. The resulting analysis aims to reveal what
aspects of pausing most effect listener perceptions of comprehensibility with possible
applications for the EFL/ESL classroom.
CHAPTER 2. LITERATURE REVIEW

2.1 Measures of Listeners’ Perceptions

Comprehensibility is not an isolated measure of listeners’ perceptions, but rather is related to accentedness and intelligibility (Munro & Derwing, 1995). Accentedness is defined as the “extent to which an L2 learner’s speech is perceived to differ from native speaker norms” (Kang, 2010, p. 302). This measure of speech is certainly important as investigations of workplace discrimination and general bias show (Munro & Derwing, 1995; Sato, 1991). However, accent reduction is also known to “be an unrealistic goal for teachers to set for their students” as native-like pronunciation is often unachievable (Pica, 1994, p. 73).

In light of these two seemingly conflicting viewpoints, it is pertinent to remember that accentedness does not necessarily inhibit understanding. A speaker can have a strong accent but still be understood. “A strong foreign accent does not necessarily cause L2 speech to be low in comprehensibility or intelligibility” (Munro & Derwing, 1995, p. 92). This finding has lent support to the idea that L2 students should be allowed to retain their accents, perhaps as a part of their identity, and the educational focus of pronunciation has thus shifted from accent reduction to comprehensibility and intelligibility (Isaacs & Trofimovich, 2012; Levis, 2005; Munro & Derwing, 1995; Munro & Derwing, 2011).

The remaining two terms, comprehensibility and intelligibility, are sometimes used interchangeably in the literature. Hahn (2004) does not make a marked distinction between the two terms, while Isaacs and Trofimovich (2012) draw a clear distinction. “Intelligibility is defined as listeners’ actual understanding of L2 speech. It is most often measured by examining listeners’ accuracy of orthographic transcriptions of L2 speech […] Comprehensibility is typically defined as listeners’ perceptions of understanding and is
measured through listeners’ scalar ratings of how easily they understand speech” (Isaacs & Trofimovich, 2012, p. 477). This study will adopt this dichotomy. Intelligibility is a measure of listener accuracy in decoding words, whereas comprehensibility scores capture how much work listeners feel it takes them to create understanding. In research, intelligibility scores are often very high (Munro & Derwing, 1995; Derwing & Munro, 1997), and provide certain insights into listener’s perceptions. Comprehensibility, however, allows researchers to tease apart which aspects of speech cause listeners to work harder to process speech and should therefore be the focus of second language classroom instruction. Comprehensibility therefore is the focus of this study.

Hahn (2004) showed that primary stress affects comprehensibility, and additional research has shown that comprehensibility is influenced by pronunciation, grammaticality, familiarity with the topic, familiarity with the speaker, and familiarity with the speaker’s accent (Anderson-Hsieh et al., 1992; Derwing & Munro, 1997; Gass & Varonis, 1984; Varonis & Gass, 1982; Wingstedt & Schulman, 1984). The principle instrument for measuring comprehensibility was designed by Derwing and Munro for a series of studies in the 1990s (Munro & Derwing, 1995; Derwing & Munro, 1997; Munro & Derwing, 1998). They devised a nine-point Likert scale with 1 as “extremely easy to understand” and 9 as “impossible to understand.” While they have made some modifications, such as changing the wording to “very easy to understand” and “very hard to understand” for later studies, this instrument has been used consistently in pronunciation research for the past twenty years.

Kang (2010) addressed reliability issues and expanded upon Derwing and Munro’s comprehensibility instrument by creating five 7-point Likert scale items. First, she reviewed the literature on speech constructs in comprehensibility and intelligibility research and then
surveyed several descriptors with meanings similar to ‘comprehensible.’ She piloted a number of items and specifically asked listeners if they perceived the descriptors in a similar manner. She also threw out items that contributed negatively to the overall reliability of the scores (O. Kang, personal communication, February 24, 2015). The five descriptors in Figure 1 constitute her final instrument and have an internal consistency of 0.94.

<table>
<thead>
<tr>
<th>The speaker to whom I just listened…</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>was easy to understand <strong>/</strong>/<strong>/</strong>/<strong>/</strong>/ was hard to understand</td>
<td></td>
</tr>
<tr>
<td>was incomprehensible <strong>/</strong>/<strong>/</strong>/<strong>/</strong>/ was highly comprehensible</td>
<td></td>
</tr>
<tr>
<td>was unclear <strong>/</strong>/<strong>/</strong>/<strong>/</strong>/ was clear</td>
<td></td>
</tr>
<tr>
<td>required little effort to understand <strong>/</strong>/<strong>/</strong>/<strong>/</strong>/ required lots of effort to understand</td>
<td></td>
</tr>
<tr>
<td>made it simple to grasp the meaning <strong>/</strong>/<strong>/</strong>/<strong>/</strong>/ made it difficult to grasp the meaning</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1. Comprehensibility instrument from Kang (2010).*

This study will adopt Kang’s five-item instrument and look at the correlation between the first item, Derwing and Munro’s original descriptor, and the remaining four items in order to verify the strength of the instrument.

There is one more perceptual measure that may be closely related to comprehensibility, and it is that of fluency. “Fluency is viewed as a kind of ‘smoothness’ or continuity of speech,” (Koponen & Riggenbach, 2000, p. 8). Just like comprehensibility, listeners seem to have a commonsense understanding of the term and therefore, features of speech can be correlated with perceptual scores to derive an understanding of the term (Lennon, 1990). Definitions of fluency often fall into two broad categories: those concerned with global proficiency and those defining fluency as one aspect of speech production (Koponen & Riggenbach, 2000). The first meaning looks at native speaker production and deals with terms like *eloquence, wit, and social appropriateness* (Fillmore, 1979). This study,
however, will adopt the second meaning, which refers to the automaticity and speed of the speech (Cucchiarini, Strik & Boves, 2002). Measures of fluency are often closely tied to pausing (Griffiths, 1991); therefore, assessment of fluency will be included in this investigation to examine how pausing affects fluency and whether fluency ratings are correlated with comprehensibility.

2.2 Pausing

Pausing is a complex phenomenon meriting its own field, pausology. Pausology has been the focus of L1 investigations since the mid-1950s, but has only been the focus of L2 studies since the 1980s. Griffiths (1991) points to the questionable methodology of early L2 pausology studies as a reason for pausology’s slow acceptance into SLA research, but strong studies can be found.

Riggenbach (1991) began her investigation of pausing and fluency by defining different types of pauses according to their length as can be seen in Table 1. While this classification will not be adopted for this study, it shows a clear example of procedures and gives insight into how pauses of different length function differently.

Table 1. Classification of pausing according to length (Riggenbach, 1991)

<table>
<thead>
<tr>
<th>Type</th>
<th>Length</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micropause</td>
<td>less than .2 seconds</td>
<td>Occur frequently in NS speech</td>
</tr>
<tr>
<td>Hesitation</td>
<td>.3 to .4 seconds</td>
<td></td>
</tr>
<tr>
<td>Pause</td>
<td>.5 seconds or greater</td>
<td>Occur in NNS speech, often over 3 seconds in length, thought to a result of language processing</td>
</tr>
</tbody>
</table>

Another approach to pausological research is to define a minimum cut-off length for analyzing pauses. There is no consensus in the L2 pausology research, but two main cut-offs have been used with consideration to how the data is being analyzed and why. Isaacs and Trofimovich (2012) and Riggenbach (1991) adopted 400 microseconds (or .4 seconds) as the
cut-off for pauses on the grounds that pauses of shorter length are a natural part of NS speech. Griffiths (1991) and Kang (2010) adopted 100 microseconds as the cut-off, drawing on L1 pausological research.

This study adopted a .1 second cut-off because the speech samples utilized were not collected from spontaneous, L2 utterances. Rather, the speech sample created endeavored to emulate L1 speech, and in this way, the location of micropauses was a salient issue.

Anderson-Hsieh and Venkatagiri (1994), who also adopted the .1 second cut-off, studied pausing in the speech of Chinese ESL speakers and found results consistent with Adams (1979), that non-native speakers paused more frequently than native speakers, paused at inappropriate places, and paused for a longer amount of time. These three aspects, pause distribution/placement, pause frequency, and pause duration are key aspects of pausing identified by Griffiths (1991) and therefore have been chosen as the three conditions to isolate in this study.

2.3 Perceptions of Pausing

Rather than dedicated pausology research, pausing is often one feature of many investigated in studies of comprehensibility and fluency. For example, Isaacs and Trofimovich (2012) investigated 19 features in phonology, fluency (temporal aspects), grammar, lexis and discourse. They found that irregular pause placement (pause error ratio), total filled pauses and total unfilled pauses were significantly correlated with comprehensibility. However because the correlations for the pause features (-0.58, -0.45, and -0.32 respectively) were lower than those of lexis (0.78), intonation (-0.76), mean length of run (0.71), story breadth (0.71) and grammar (-0.63), pausing was not considered a salient feature when creating a grading rubric.
Cucchiarini et al. (2002) investigated seven temporal aspects of speech including mean length of silent pauses and number of silent pauses per minute and found that while less fluent speakers had longer pauses and paused more frequently, increased frequency of pauses was more relevant to perceived fluency than the length of the pauses.

Kang (2010) studied both accentedness and comprehensibility and looked at 12 features in fluency, stress and pitch. She found that longer, inappropriately placed pauses correlated with accentedness while frequency of pauses (phonation-time ratio) and the mean length of filled pauses had the highest correlations to comprehensibility.

Kormos and Dénes (2004) investigated accuracy, lexical diversity and 10 temporal measures related to fluency and found that phonation time ratio (the frequency of pauses) was a good predictor of fluency ratings but was not as strong as mean length of run and speech rate. The effect of the mean length of pauses differed with regards to the rater, and the number of filled and unfilled pauses had no influence on perceptions of fluency.

While these studies provide a wealth of data, it is difficult to arrive at a clear understanding of the effects of pausing. Studies focusing discreetly on pausing are clearer. Lennon (1984) reported that the location of pauses within clauses and not at clause junctures could reliably discriminate between fluent and non-fluent speakers. Tavakoli (2011) supported this finding. He recorded 40 native and 40 non-native speakers in a narrative task and analyzed the recordings for the number of pauses in the middle of clauses and at the end of clauses and the total amount of silence in the middle of clauses and at the end of clauses. While non-native speakers did pause more frequently and had longer pauses than native speakers, these differences were not significant. What distinguished the two groups was non-native speakers’ frequent pauses in the middle of clauses rather than at the end. Tavakoli
hypothesized that the online-nature of speaking and the need for planning caused these differences in native and non-native speaker pausing. Both of these pausological studies, Lennon (1984) and Tavakoli (2011), emphasized the importance of placement over the other pausing features of length and frequency.

2.4 The Current Study

Looking at the intersection of comprehensibility and pausing, this study examines three research questions:

1) How do frequency, length, and placement of pausing affect ratings of comprehensibility? Following the findings of Lennon (1984) and Tavakoli (2011), we hypothesize that pause placement will be the most salient feature affecting comprehensibility. Following pause placement, pause frequency will affect perceptions, and finally pause length. This is in accordance with the findings of Cucchiarini et al. (2002).

2) How do the effects of pausing on comprehensibility compare with the effects of pausing on ratings of fluency? We hypothesize that increased frequency of pauses and increased length of pauses will show the smallest correlations between comprehensibility and fluency. This is because both frequency and length influence rate of speed and phonation time ratio, which are traditionally categorized as concepts of fluency. If indeed the comprehensibility and fluency ratings are measuring different constructs, they should show low correlations for these two pause treatments.

3) How well do the additional descriptors introduced by Kang (2010) capture the original “easy/hard to understand” descriptor in the creation of a more robust comprehensibility instrument? We hypothesize that given Kang’s systematic creation of her instrument, that all of the descriptors will highly correlate with the primary descriptor. Of
the four additional descriptors, we predict that “clear/unclear” will have the lowest correlation to “easy/hard to understand” because it is the furthest from “easy/hard to understand” in meaning.
CHAPTER 3. METHODOLOGY

To investigate the effects of pausing on comprehensibility and fluency, four classes of university underclassmen were asked to listen to each of four different pause treatments and rate them for comprehensibility and fluency on 7-point Likert scales. The original speech sample (Version A) was recorded by a female Korean national with advanced level English proficiency, and then this recording was modified according to pause placement (Version B), pause frequency (Version C) and pause length (Version D). Next, recordings were created to make sure that each class heard the treatments in different orders. This was a departure from Hahn’s (2004) methodology in which each group listened only to one version. Instead, in this study each group listened to each version, albeit in different orders. This accounted for the non-randomized nature of the classes. Finally the data was collected and analyzed.

3.1 Materials

3.1.1 Participants

Forty-three participants were recruited from four first-year composition courses at a public Midwestern university. This population was chosen so that this study would be comparable with previous comprehensibility and fluency research. Second, this population is relatively homogeneous with regards to age, cultural background, life experience, exposure to foreign accents, and travel abroad. Third, all university undergraduates must take the first-year composition sequence, ensuring a wide cross-section of the university population and the representation of a variety of majors. This population was solicited only after receiving exempt status from the Institutional Review Board (IRB ID 14-571).

Of the 43 participants, 37 (86%) were freshman, 3 were sophomores and 3 were transfer students. The majority of the participants were male, in-state residents; however,
this reflects the actual make-up of the university as seen in Table 2. All spoke English as their first language, and one participant self-identified as bilingual. Nearly all of the respondents had had previous experience with non-native speakers: 65% had been abroad, 91% had had at least one international teaching assistant as an instructor, and 72% self-identified as having had regular contact with a non-native speaker for more than a month.

Six participants indicated that they had read or heard about the topic of the lecture before, and while this is known to affect comprehensibility ratings (Gass & Varonis, 1984), these responses were retained. Of the 43 participants, distribution was similar across the four classes as seen in Table 3.

Table 2.
Demographics of the participants and the university (Iowa State University Office of the Registrar, 2014)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Participants</th>
<th>University Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>74%</td>
<td>57%</td>
</tr>
<tr>
<td>Female</td>
<td>26%</td>
<td>43%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Origin/Home</th>
<th>Participants</th>
<th>University Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-state</td>
<td>49%</td>
<td>64%</td>
</tr>
<tr>
<td>Out-of-state</td>
<td>46%</td>
<td>28%</td>
</tr>
<tr>
<td>International</td>
<td>5%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Table 3.
Distribution of participants across classes

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>12</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>

In addition to these 43 participants, there were 34 other respondents, but their data was excluded. Five self-identified as having hearing difficulties such as hearing loss, buzzing in the ears, or noise-induced hearing-loss. Excluding these data points is consistent with the work of Munro and Derwing (1995; 1997; 1998). There was also one student who
despite instructions was playing on her personal computer, so her data was excluded. The remaining 28 respondents were non-native speakers, and these responses were excluded (Munro & Derwing, 1998). However, including the responses of non-native speakers should be considered for future research, not only for reasons of social justice, but also to accurately reflect the populations of students in the American university context (Kormos & Dénes, 2004).

3.1.2 Speech Sample

In keeping with the university context, an authentic transcript of an academic lecture was used as the basis for the speech samples which the 43 participants rated. The text came from the Michigan Corpus of Academic Spoken English (MICASE), a roughly 2 million word corpus of naturally occurring collegiate discourse. A variety of entry-level, informative lectures from the humanities (history and the arts) were surveyed by selecting the Lecture-large speech event type and the Humanities and Arts academic division within the MICASE online interface. The choice of humanities over sciences was purely preferential on the part of the researcher and is allowable given that content is a controlled variable in this study.

After reviewing a number of lecture transcripts, the researcher chose a lecture about women’s protective legislation at the turn of the century from a course called the History of the American Family. This lecture was chosen for its avoidance of field-specific jargon. A section of roughly 500 words was isolated in order to yield a recording of 2-3 minutes. Choosing a discourse length text, as opposed to sentence-length samples, is in accordance with Levis and Pickering (2004) which advocated for the use of discourse level speech samples due to the influence of surrounding information on the pronunciation of a text. After selection, the text was modified to reduce unusual wording on the part of the original speaker.
(others who were supportive of bolstering labor’s position and a lucrative and fairly decent-paying niche) and to eliminate phrases or clauses that might prove unwieldy upon recitation (and the debate over protective legislation that historians have been carrying on ever since the progressive period still goes on). The modified transcript of 418 words can be found in Appendix A.

While the original text was naturally occurring spoken text, for this experiment the recording was read, rehearsed speech. This is a recognized limitation of the methodology; however, it should be noted that preparing and reciting a lecture is similar to read, rehearsed speech and is thus close enough in style and execution for the context of this study (Hahn, 1999).

3.1.3 Recordings

The normal recording (Version A) was read by a female Korean national with a high level of oral English proficiency and many years of experience in the collegiate L2 environment. A non-native speaker was chosen over a native speaker so that the results could be applied to the EFL/ESL context. To explain, non-native speakers inherently have accents, and we wanted to represent that, albeit as a controlled variable. If a native speaker had created the recording, the results would not necessarily be generalizable to the second-language context, but using a non-native speaker allowed for the representation of accent in the experimental construct. Further, a non-native speaker maintained authenticity, as the pause treatments attempted to reflect common features of non-native speech. This specific speaker was chosen for her clear segmental production. She also spoke confidently without mumbling. All of these indicate her very high ability.
The reader was given the script in advance of recording so that she could investigate any unfamiliar words (*contingent*) and become comfortable with the meaning. On the designated day, she was recorded multiple times over a period of one hour ensuring that ambient noise and vocal quality would be consistent across the recordings. The sample was recorded with Audacity (Mazzoni & Dannenberg, 2000), a freeware audio editor, on the built-in microphone of a 2011 MacBook Pro. The recording took place in the university center for new media which was designed to reduce echo and extraneous noise. After review, the recording with the highest quality in terms of background noise and clarity as well as consistent pacing (fewest hesitations and abrupt pauses) was selected.

To confirm that the quality of the recording was native-like with regards to pausing, three native speakers of English were asked to record the sample in the same manner as the non-native speaker, and the recordings were compared for length, number, and placement of pauses. The data for this analysis is found in Appendix B. Overall, the non-native recording was longer and had more pauses than the native speakers. The native speakers averaged 2 minutes and 25 seconds for the recording, and the non-native speaker was 2 minutes and 48 seconds. The native speakers averaged 52 total pauses, and the non-native speaker had 58. For any non-native speaker, this is an expected outcome (Tavakoli, 2011), but does not preclude the “native-like” nature of their pausing. Deviation within native speaker’s pausing must also be taken into consideration.

With regards to the total number of pauses, the native speakers deviated plus or minus four pauses from their average. The non-native speaker only had two more pauses than the highest native speaker (56 pauses), and as such was still quite close to the range of the native speakers. Specifically looking at length, 44% of non-native speaker’s pauses were within the
range of length of the native speaker production; 82% were within 200 milliseconds of the native speaker range. This is acceptable following the analysis of the three native speakers. They only agreed on length within 200 milliseconds 75% of the time. The last factor for analysis was pause placement, and the non-native speaker sample matched pause placement with all three native speakers or two of three native speakers 79% of the time. If we look at non-native speaker placement matching even just one native speaker, the percentage rises to 92% of the time. Clearly, there is a large amount of agreement on placement. The amount of agreement is further supported when we look at original pauses, or pauses shared by no other speaker. Among the native speakers, the first speaker had eight unique pauses shared with no other native speaker, and the other two each had four unique pauses. The non-native speaker, however, only had three unique pauses shared by no other speaker. All of this data on frequency, length and placement reveals that pausing is variable, even among native speakers. However, this non-native speaker recording, despite its slightly longer length and increased number of pauses, was within the range of native speaker variation. Therefore, it was determined to keep the non-native speaker recording as it was, without modification, not only because the recording was clearly close to native-like production but also to maintain the authenticity of the recording. This final non-native speaker control transcript with pauses marked may be found in Appendix C.

The next step was to create the experimental recordings; however, first the lecture needed to be subdivided into sections. This was a departure from Hahn’s (2004) original methodology. In Hahn (2004), each of the participants was randomly placed in a group and listened to only one of the variations. However, in this experiment, the participants came from pre-determined classes, and as such could not be randomly assigned to groups. In order
to control for the effects of class, it was decided that each of the respondents would listen to each of the different variations, and each class would hear the variations in different orders to control for ordering effects. This decision not only mitigated the grouping of the participants, but also increased the amount of data collected. This change in the methodology was also facilitated by the fact that respondents can make reliable judgments on comprehensibility with only 20 seconds of speech (Derwing & Munro, 1997). Therefore, it was decided that respondents would listen to sections of the lecture, 20 to 30 seconds at a time. In this way they could rate all four different pause variations (normal and three experimental) and still hear the pause variations within the larger discourse context of the section.

The lecture was divided into six sections. The first section was from the normal recording and acted as a baseline. This introductory section also allowed respondents to become familiar with the rating instrument. The last section was another normal section, but it more accurately served as a catch-all for the tail-end of the lecture, after the other four sections had been equalized for the number of pauses. In choosing to divide the lecture, both semantic breaks (lecture ideas and paragraph markings) and the number of pauses were taken into account. For the most part, the lecture was divided along semantic breaks, with the exception of the break between Section 2 and Section 3. This break helped equalize the number of pauses per section as seen in Table 4. The six sections can be seen in the transcript in Appendix C.

Table 4.
*Distribution of pauses across the sections of the lecture*

<table>
<thead>
<tr>
<th></th>
<th>Section 1</th>
<th>Section 2</th>
<th>Section 3</th>
<th>Section 4</th>
<th>Section 5</th>
<th>Section 6</th>
</tr>
</thead>
<tbody>
<tr>
<td># of pauses</td>
<td>X</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>X</td>
</tr>
</tbody>
</table>

Following the finalization of the transcript, three experimental recordings were created. These were Version B with irregularly placed pauses, Version C with frequent
pauses, and Version D with longer pauses. For all three experimental versions, 2/3 of the pauses in each section were modified to the desired pause type, leaving 1/3 of the section in its original form. This was for the sake of cohesion and naturalness and follows the precedent set by Hahn (1999). To determine which pauses to modify, all 39 pauses across the four sections were listed in an Excel spreadsheet. The first pause of each section remained unmodified to allow the listeners some time to get used to the speaker before modifications were introduced. Then, the next two pauses were marked for modification. This pattern of one unmodified followed by two modified was repeated until the section was complete. This ensured that the distribution of modified pauses was balanced across any given section. This meant that there were six modified pauses per section.

To create Version B with irregularly placed pauses, pauses were moved from the ends of phrases and clauses to within phrases as well as to between an article and a noun (Kang 2010). The pauses that were moved retained their original length. For example, *Factory owners were forced to establish women’s bathrooms* (.5). *Women were prevented from doing certain kinds of labor* (.3), was modified to: *Factory owners were forced to establish* (.5) *women’s bathrooms. Women were prevented from* (.3) *doing certain kinds of labor.* As much as possible, the researcher tried to keep the new placement of the pauses close to the old pause placements in order to maintain the distributional balance of modified pauses across the section.

In Version C, additional pauses, six per section, were placed within the recording in accordance with appropriate thought groups. Inserted pauses were .4 seconds in length. 400 milliseconds is an average pause length, and as indicated by Riggenbach’s (1991) taxonomy, falls in the middle of the pause spectrum. To determine appropriate placement, the native
speaker samples were used, and any placement that a native speaker used was validated as appropriate. After pauses following native speaker placement were inserted, the remaining modified pauses were placed after clauses, after discourse markers such as also, however and then, and before prepositional phrases.

For Version D, 2/3 of the pauses were doubled in length. A pause of .7 seconds was doubled to 1.4 seconds and a pause of .3 seconds was doubled to .6 seconds. The three experimental pause recordings can be found in Appendix D, E, and F.

In creating the experimental recordings, a number of challenges were faced. First, using computer created silence to insert the pauses was jarring and unnatural. Even using ambient silence from the original recording was ineffective. To make the pauses sound natural, silence which included background noise or inhalation sounds were used. Background noise included rustling or movement (not computer sounds like clicking) and generally looked like little crumbs on the waveform. This background noise is generally imperceptible unless you are listening for it, and is more noticeable in its absence than when it is present. The use of inhalation, small breaths to prepare for speech, was especially important in the creation of the irregular placement recording. Ambient pauses out of place were quite jarring, but inserting an inhalation made the pause acceptable to the listener. The breath does so much to prepare the listener, but we are generally unaware of its importance.

The second difficulty was in placement. On multiple occasions, a pause could not be added or moved to a desired location because in the original recording the words were elided or part of the same waveform. Also, the intonation of a word or its short length made a pause following the word feel awkward. The word became clipped and unnatural. To deal with
these issues, secondary and tertiary placements were considered when moving and adding pauses. Above all, the goal was to maintain the naturalness of the speech.

After the three experimental recordings were created they were shared with nine ESL teachers (three teachers for each sample) to rate as either a) possible that a non-native speaker would produce or b) unnatural. The three teachers who rated Version C with increased pause frequency all rated the sample as possible that a non-native speaker would produce. Of the three teachers who rated Version D with longer pauses, two rated it as natural and one rated it as unnatural. This teacher explained that it sounded like read speech instead of spontaneous speech, but did not mention anything about the prosody. Therefore, Version D was also validated on its naturalness. Version B, however, was jarring for two of the three teachers. With each of these teachers, the researcher sat down and asked the teachers what they found unnatural about the recordings. The teachers identified places in which pauses between sentences had been removed and the new sentence began with a paratone. A paratone is the use of a high pitch to mark the beginning of a paragraph or new idea (Levis & Pickering, 2004). The preceding sentence ends with declining pitch, and the paratone marks a topic shift for the listener. The total absence of pausing at these junctures made the sentences unnatural, so the researcher reinserted .2 seconds of pause in five different locations to mitigate the paratones. While this revision was less than ideal, it was also a natural consequence of manipulating one control recording. Reinserting a micropause and creating a more natural recording was a good compromise in order to maintain accent, intonation, and speed across all of the recordings. In this way, Versions B, C, and D were completed.
The final step before data collection was to actually create the recordings that would be used in the classrooms. These recordings needed to present each of the four variations (normal and three experimental) in Sections 2-5 in a randomized order. This was achieved through the creation of a Latin square. A Latin square looks somewhat like a Sudoku grid with each number (representing a treatment) in each column and row without overlap. A variety of different Latin squares can be created, but for optimal randomization, no treatment can be followed by any other treatment more than once. To illustrate, look at the two Latin squares in Figure 2.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>A</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>A</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>D</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Latin Square #1
Latin Square #2

*Figure 2. Two possible Latin squares. Although Latin squares are usually represented with numbers, letters were used here to represent the different treatments.*

Both Latin squares conform to the rules of Latin squares, with one letter in each column and row. However, when looking at Latin square #1, Version A is followed by Version B twice and Version D once. Therefore, this Latin square may have an ordering effect. In Latin square #2, however, Version A is followed once each by Versions B, C, and D. This randomization is more balanced and will show fewer ordering effects. Therefore, Latin square #2 was used in this experiment and was the basis for the finalized class recordings which can be found in Table 5.
In addition to the randomized treatments and baseline sections, each class recording began with directions. Each section was announced and at the completion of each section “Stop and rate” prompted the participants. This facilitated smooth data collection.

Table 5.
*Class recordings with randomized distribution of the four sample recordings*

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section 1 – Normal Baseline</strong></td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td><strong>Section 2</strong></td>
<td>Normal</td>
<td>Irregular Placement</td>
<td>Increased Frequency</td>
<td>Length</td>
</tr>
<tr>
<td><strong>Section 3</strong></td>
<td>Increased Frequency</td>
<td>Normal</td>
<td>Length</td>
<td>Irregular Placement</td>
</tr>
<tr>
<td><strong>Section 4</strong></td>
<td>Irregular Placement</td>
<td>Length</td>
<td>Normal</td>
<td>Increased Frequency</td>
</tr>
<tr>
<td><strong>Section 5</strong></td>
<td>Length</td>
<td>Increased Frequency</td>
<td>Irregular Placement</td>
<td>Normal</td>
</tr>
<tr>
<td><strong>Section 6 – Second Baseline</strong></td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

3.2 Procedures

The researcher visited each classroom, described the study, and invited the students to participate. While students had the option not to participate, all students agreed to fill out the survey. After gaining permission, the researcher distributed the two-page questionnaire, and students filled out the first page with demographic questions and questions about familiarity with non-native speakers, familiarity with the topic, and hearing impairment. When all participants had completed the background information, the researcher began to play the
appropriate class recording. After each section, the students filled-out the five-item Kang comprehensibility instrument as well as one additional question about fluency.

- fluent ___/___/___/___/___/___/___ not fluent

Upon conclusion of the sixth section, the researcher collected the responses and thanked the students for their participation. Finally, the data was coded and recorded in an Excel spreadsheet. The questionnaire can be found in Appendix G.

3.3 Analysis

In choosing to run a statistical test, the researcher considered the model that the data would be compared against. A simple model looks at the primary variable under investigation and compares the data to a null model. This type of test usually assumes that any other variables have been accounted for or balanced out in another way. A more complex model considers many variables and compares the data against this construct, rather than a null model.

In this experiment there was one primary independent variable of pausing, but there were also two other independent variables of class and order. The dependent variable was the comprehensibility rating. Given this set-up, the researcher chose a more complex model for statistical analysis, and a three-way ANOVA test was run. The three-way means that all three independent variables were accounted for in the construction of the model. These three factors were treatment, order, and class. A fourth factor was also included in the analysis, that of student, which was nested within class. ANOVA was chosen because the independent variables had multiple levels. Pause treatment had four levels (normal, placement, frequency, and length) as did the variables of order and class. ANOVA is the appropriate test when the
independent variables have different levels and the resulting dependent variable is continuous with a normal distribution (Larson-Hall, 2010).

The data set met these requirements, and the 172 data points were sufficiently above the minimum sample size of 30 (Larson-Hall, 2010). The only violation of the ANOVA assumptions was that the dependent variable should be interval rather than ordinal. The comprehensibility rating was ordinal because only the first response (“easy/hard to understand”) was used for analysis. Because the remaining four items were under investigation, it was felt that only looking at the first response was the stronger choice. In future research, Kang’s instrument may be useful as an average of five items creating an interval rating.

To conduct the test, the researcher went to the university statistics help desk and with their help set up the test in JMP, statistical software produced by the SAS Institute. Tukey post-hoc tests were also run for the independent variables of order and treatment.

To answer the second research question concerning the similarity between comprehensibility and fluency, the researcher conducted Spearman correlation tests between the first comprehensibility rating (“easy/hard to understand”) and the fluency rating. Spearman is the appropriate correlation test because both variables are ordinal (Bachman, 2004). Five different tests were run, the first comparing all comprehensibility ratings with all fluency ratings, and then four more tests one for each of the different pause treatments. These additional correlations had the potential to show if different pause factors were related to the ratings of comprehensibility or fluency.

The final research question investigated the strength of Kang’s five-item instrument as compared to the original one-item “easy/hard to understand” response. To look at this
issue, the averaged comprehensibility rating produced by Kang’s instrument was compared to the first item response. This was done for the aggregate responses, as well as for each of the four pause treatments separately. Following this, the first response item was compared to each successive item, for both the entire data set and for each individual pause treatment. This helped isolate if any of the pause treatments caused the items to capture different information than the original “easy/hard to understand.” For all of these correlations tests Spearman was used. This is appropriate even for the tests which compared the averaged rating to the first item response because Spearman is also used when one variable is ordinal and the other is interval (Bachman, 2004).

For the Spearman rho rank correlation tests, any value larger than 0.7 was considered strong, and any value smaller than 0.3 was considered weak. The p-value was set at .05. All of the Spearman calculations were completed on R, statistical freeware (University of Auckland Statistics Department, 1997).
CHAPTER 4. RESULTS

4.1 Effects of Pause Treatment on Comprehensibility

The first research question asked if pause treatment had any significant effect on comprehensibility ratings. A three-way ANOVA blocking for class, student nested within class, order and pause treatment was run, and the result was statistically significant with $F(48, 123) = 5.895$, $p$-value $< .0001$. This means that the model with all of the factors incorporated accounted for a significant portion of the variance between the responses. The effect tests reported in Table 6 show how much each variable accounted for variance in the data. The individual students accounted for the most of the variance with a DF (degrees of freedom) value of 39. This means that more than any other variable the individual students caused most of the difference in the data. For example, one student may have had ratings of 1, 2, 2, and 1 for the four treatments, but another student may have had ratings of 5, 6, 7, and 6. These are quite different, but the statistical model accounts for this difference and reports that individuals are responsible for big differences such as these. The other three factors, class, order and pause treatment, equally explained the remaining variance within the data with a DF of 3. All factors were statistically significant.

Table 6.
Effect tests of the three-way ANOVA

<table>
<thead>
<tr>
<th>Factor/Source</th>
<th>DF</th>
<th>F Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student[Class]</td>
<td>39</td>
<td>5.745</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Class</td>
<td>3</td>
<td>5.414</td>
<td>0.0016</td>
</tr>
<tr>
<td>Order</td>
<td>3</td>
<td>7.340</td>
<td>0.0001</td>
</tr>
<tr>
<td>Pause Treatment</td>
<td>3</td>
<td>6.274</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Looking at the effects of treatment, the averages and standard deviations by class are presented in Table 7. Irregular placement was rated as the hardest to understand with an
average of 4.628 across all of the classes. This was followed by increased frequency with an average of 4.326. The comprehensibility ratings of the normal recording and the recording with long pauses dropped significantly compared to placement and frequency with averages of 3.674 and 3.860 respectively.

Table 7.

Means and standard deviations for the different treatments by class

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Irregular Placement</th>
<th>Increased Frequency</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>3.417</td>
<td>4.917</td>
<td>4.083</td>
<td>4.417</td>
</tr>
<tr>
<td></td>
<td>2.021</td>
<td>2.151</td>
<td>1.975</td>
<td>1.776</td>
</tr>
<tr>
<td>Class 2</td>
<td>2.778</td>
<td>3.333</td>
<td>4.111</td>
<td>3.778</td>
</tr>
<tr>
<td></td>
<td>1.302</td>
<td>1.871</td>
<td>1.764</td>
<td>1.856</td>
</tr>
<tr>
<td>Class 3</td>
<td>4.700</td>
<td>5.200</td>
<td>4.100</td>
<td>3.800</td>
</tr>
<tr>
<td></td>
<td>1.160</td>
<td>1.229</td>
<td>1.287</td>
<td>1.549</td>
</tr>
<tr>
<td>Class 4</td>
<td>3.750</td>
<td>4.833</td>
<td>4.917</td>
<td>3.545</td>
</tr>
<tr>
<td></td>
<td>1.545</td>
<td>1.467</td>
<td>0.996</td>
<td>1.084</td>
</tr>
<tr>
<td>Total</td>
<td>3.674</td>
<td>4.628</td>
<td>4.326</td>
<td>3.860</td>
</tr>
<tr>
<td></td>
<td>1.658</td>
<td>1.800</td>
<td>1.539</td>
<td>1.567</td>
</tr>
</tbody>
</table>

Note. The top bold number is the mean and the bottom number is the standard deviation. Comprehensibility is on a scale from 1 to 7, with 1 as “easy to understand” and 7 as “hard to understand.”

The Tukey post-hoc test further confirmed that differences for irregular placement and frequency were significant. While both were significantly different than the normal treatment, length was not. Within the treatments, irregular placement was significantly different than length, but increased frequency was not different than irregular placement, nor was increased frequency different than length. This data with means and p-values can be found in Table 8.

Participants found irregular placement the most difficult to understand. Increased frequency was the next most difficult to comprehend and was as salient as irregular placement in its effect on comprehensibility. The recording with longer pauses, however,
was rated similarly to the normal recording. The fact that length was found to differ significantly with irregular placement as opposed to increased frequency indicates that the irregularly placed pauses were more difficult to listen to than the recording with the increased number of pauses.

It is also interesting to note that there was a significant effect for order. The design of the experiment attempted to account for ordering effects by ensuring that each class listened to the treatments in different orders, but the results revealed that there was still an effect.

When the students rated, Section 4 at the top of the second page was significantly different from Section 2 (mean difference = 0.990, CI = 1.602, 0.378, p = 0.0003) and Section 5 was
significantly different from Section 2 (mean difference = 0.784, CI = 1.394, 0.174, p = 0.0059). Possible reasons for this will be considered in the discussion section.

4.2 Comprehensibility vs. Fluency

The second research question asked how the ratings for comprehensibility compared with those of fluency. When ratings were compared for all 172 data points, comprehensibility and fluency were found to have a strong correlation at $r = 0.722$ ($p < .001$). This high correlation means that in the minds of students the ratings of comprehensibility and fluency were closely related.

The correlations between comprehensibility and fluency for the four different pause treatments can be seen in Table 9. As expected, comprehensibility and fluency were most closely related when students listened to the normal recording ($r = 0.816$, $p < .001$). This does not tell us much about what makes something comprehensible or fluent, only that when a speaker is highly comprehensible or fluent, it is easy for native speakers to identify, and they are confident in their identification. On the other end, length had the lowest correlation for comprehensibility and fluency ($r = 0.612$, $p < .001$). This indicates that with respect to length, the fluency rating is capturing something different from the comprehensibility rating.

Table 9.

<table>
<thead>
<tr>
<th></th>
<th>Spearman correlations for comprehensibility and fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Data</td>
<td>$r = 0.722$</td>
</tr>
<tr>
<td>Normal</td>
<td>$r = 0.816$</td>
</tr>
<tr>
<td>Irregular Placement</td>
<td>$r = 0.702$</td>
</tr>
<tr>
<td>Increased Frequency</td>
<td>$r = 0.749$</td>
</tr>
<tr>
<td>Length</td>
<td>$r = 0.612$</td>
</tr>
</tbody>
</table>

Note. All correlations were found to be statistically significant at the $p < .001$ level.
4.3 Kang’s Modified Instrument

The final research question investigated the strength of Kang’s instrument and asked if the four items captured the same idea as the original “easy/hard to understand” item. First, correlations between the first item and the five averaged items were calculated. The Spearman correlation for all of the data was strong with $r = 0.924$ ($p < .001$). For the normal recording, the first item and the averaged items had a correlation of $r = 0.966$. For irregular placement it was $r = 0.905$, for increased frequency it was $r = 0.893$, and for length it was $0.924$. All were statistically significant at $p < .001$. This means that there was a strong correlation between the original item and the five-item instrument. This result is not surprising considering that the first item is included in the average of the five items, naturally leading to higher correlations. Further, the process of averaging the five items mitigates any outliers, again leading to higher correlations.

More interesting observations appear when we look at the correlation between the first item and the four other items, as seen in Table 10. Only three of the 20 $r$-values are below 0.7, and even those are close to 0.7 at 0.688 and 0.693. The values below 0.7 have been shaded in grey.

Of all of the descriptors “clear/unclear” most closely captured the meaning of “easy/hard to understand” with an $r$-value of 0.812. “Incomprehensible/comprehensible” had the weakest correlation of the four items with $r = 0.704$. 
Table 10. Spearman correlations between “easy/hard to understand” and the remaining items

<table>
<thead>
<tr>
<th>“easy/hard to understand” compared to…</th>
<th>“incomprehensible/comprehensible”</th>
<th>“clear/unclear”</th>
<th>“required little/ lots of effort”</th>
<th>“simple/difficult to grasp”</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Data</td>
<td>0.704</td>
<td>0.812</td>
<td>0.785</td>
<td>0.761</td>
</tr>
<tr>
<td>Normal</td>
<td>0.707</td>
<td>0.810</td>
<td>0.867</td>
<td>0.832</td>
</tr>
<tr>
<td>Irregular Placement</td>
<td>0.623</td>
<td>0.774</td>
<td>0.743</td>
<td>0.708</td>
</tr>
<tr>
<td>Increased Frequency</td>
<td>0.688</td>
<td>0.763</td>
<td>0.693</td>
<td>0.789</td>
</tr>
<tr>
<td>Length</td>
<td>0.804</td>
<td>0.869</td>
<td>0.847</td>
<td>0.704</td>
</tr>
</tbody>
</table>

Note. All correlations were found to be statistically significant at the p < .001 level. Values below 0.7 have been shaded in grey.
CHAPTER 5. DISCUSSION

In investigating the effects of pause treatment on comprehensibility ratings, the three-way ANOVA confirmed the hypothesis that irregular placement had the greatest effect on comprehensibility scores, followed by increased frequency. This finding is compatible with the work of Lennon (1984) and Tavakoli (2011). Lennon (1984) found that the placement of pauses within clauses discriminated between fluent and non-fluent speakers. Tavakoli (2011) likewise found that non-native speakers used more pauses within clauses than native speakers. Thus, the use of pausing within clauses rather than at clause boundaries, simulated in this experiment with irregular placement, has significant impacts for understanding and for evaluation of speaking ability. So, why is pause placement so salient? I propose that the answer lies in formulaic language.

Wood (2006) defines formulaic language as “fixed strings or chunks of words that have a range of functions and uses in speech production and communication and seem to be cognitively stored and retrieved by speakers as if they were single words” (Wood, 2006, p. 14). While there seems to be consensus about the importance of these multi-word units, there is dispute about the naming and the definition of these units. Schmitt (2000) uses the terms lexical chunk and lexical phrase to refer to these units that facilitate the processing of words. Biber, Conrad and Cortes (2004) offer the terms lexical phrases, formulas, routines, fixed expressions, prefabricated patterns, and lexical bundles. Nomenclature aside, there are two main research avenues for these multi-word units. The first, corpus linguistics, narrowly defines formulaic language with a specific length and a high frequency in a corpus. Consequently, this field prefers the terms lexical bundles, recurrent word-combinations, and n-grams (Nesi & Bastrukmen, 2006). While these computer-derived chunks are sometimes
complete units, such as on the other hand and at the same time, they are more often incomplete phrases, containing function as opposed to content words. Some examples include is one of the, want to do is, and going to have to (Biber et al., 2004). Pedagogy, fluency, and pronunciation research, however, use native speaker judgment to identify the units and tend to use the term formulaic language (Wood, 2006). This research often focuses on predictable phrases such as How are you? and Will you marry me? and idioms like beat around the bush and learn the ropes (Lindstromberg & Boers, 2008; Wood, 2006).

While these two definitions may appear incompatible, there are two recurrent ideas that appear in the literature on multi-word units, regardless of the domain. First, these units perform some communicative function in speech. Wray (2002) proposed three functional categories of formulaic language: enabling the manipulation of information, continuing the flow of speech, and signaling organization. Biber et al. (2004) within the corpus-linguistics framework established the three functional categories of stance expression, discourse organization, and referential expression. Therefore, communicative function is key to the definition of formulaic language. The second key aspect of formulaic language is that following the psycholinguistic model of language processing, it is believed that these units “are stored in and retrieved from long-term memory as if they were single lexical units” (Wood, 2006, p. 16). This would allow for faster processing especially in online activities such as speaking (Schmitt, 2000).

These two ideas taken together have implications for pausing. If the chunks are stored as a single unit in memory and have a singular function, then there should be no pausing within the unit. Dahlmann and Adolphs (2007) investigated this very question using a learner corpus and found no internal pauses for the unit I don’t know out of 47 spoken
instances. The results for *I think I* were not as strong, but still 82% of the 44 instances had no internal pausing. The unit was spoken as intact without pause. This idea of chunks of language also explains the results of this study. When the recording broke up a unit of language, as part of the irregular placement or increased frequency treatments, the native listeners could no longer decode the unit as a chunk, but had to process each word to make meaning. They were disrupted in their listening process, and this disruption then accounts for the lower comprehensibility ratings.

Tying together the work on *lexical bundles* in corpus linguistics and that of *formulaic language* in pronunciation presents some interesting ideas. While chunks identified by corpus studies are often incomplete (*and this is the, you know if you, etc.*), perhaps when complete they indicate larger thought groups. These thought groups are not interrupted with pauses and are themselves formulaic. Perhaps it is this formulaic nature of thought groups and pauses that allows for quick and efficient language processing. The very expectedness of thought groups or chunks effects language processing. This hypothesis requires further investigation, but for now this discussion will continue by reviewing current research on formulaic language.

Research has tied the acquisition of lexical chunks to higher proficiency (Boers et al., 2006; Ejzenberg, 2000; Oppenheim, 2000; Wray, 2002, cited in Lindstromberg & Boers, 2008, p. 424) and to fluency (Dechert, 1980, cited in Woods, 2006, p. 17). In fact, current research has moved beyond trying to establish the connection between formulaic language and fluency and is investigating how best to teach formulaic language (Lindstromberg & Boers, 2008) and how students use formulaic chunks in their speech (Woods, 2006). This
study offers the new understanding that the command of lexical chunks not only effects fluency but also affects comprehensibility, or the ease of understanding a speaker.

While irregular placement and frequency were indeed found to disrupt native listeners, it is interesting to note that increased length of pauses did not. Speech samples with longer pauses were processed the same as the normal recording and indeed, if we return to the data presented in Table 7, we see that both Class 3 and Class 4 rated the sample with longer pause length as easier to understand than the normal recording (Class 3: normal avg. = 4.700, length avg. = 3.800; Class 4: normal avg. = 3.750, length avg. = 3.545). This means that the longer pauses actually made the speaker easier to understand, not harder.

As teachers, we may want to interpret this finding such that slower speech is easier for native listeners to process; however, Munro and Derwing (1998) show that slower is not necessarily better. They had 20 participants, 10 Mandarin and 10 Canadian English speakers, record a narrative, once at normal speed and once at half their normal rate, and they found that listeners rated the slower narrative as more accented and less comprehensible. The intuition that slower is better is contradicted in the research. Likewise Anderson-Hsieh and Koehler (1988) found that faster is not necessarily better as faster-than-normal recordings were rated as less intelligible than normal-rate speech. Munro and Derwing (1998) proposed that the optimal speed for non-native speakers is somewhat slower than the optimal speed for native speakers. EFL/ESL students should try to match the speed of native speakers, and while they will typically fall short of this goal, the slightly slower speed will give native listeners extra time to process their accents and to understand their utterances.

What this study offers is the understanding that while non-native speakers should emulate the speed of native speakers, the placement of their pauses affects understanding. If
students pause at sentence or clausal boundaries, even if these pauses are longer, they will likely be understood. However, if students pause within phrases or clauses, this will cause disruption. In bringing this finding to our students, we first need to emphasize the importance of pauses in speaking and then help them to reframe their processing. They need to think about their utterance and produce it in one chunk. They can pause before and after the chunk without lowering comprehensibility, but pausing in the middle can cause disruption.

The importance of pause placement can be seen if we return to the original recording produced in this study. The non-native reader had longer and more frequent pauses than the native speakers, but in pause placement, she agreed with at least one native speaker on pause placement 92% of the time. This factor unconsciously set her aside as an advanced speaker of English and made her speech easy to understand and process.

Speaking as fast as native speakers may ultimately be an unobtainable goal for non-native students, similar to native accent. However, if EFL/ESL students consider what they want to say and pause appropriately, comprehensibility is a likely result.

Another finding of the three-way ANOVA was that despite a strong experimental design there was an ordering effect with significant differences for Sections 4 and Sections 5. These sections were at the top of the second page, in the latter half of the survey. Content may have affected the ratings, but a review of the lecture content did not reveal anything exceptional in either vocabulary or organization. Sections 4 and 5 discussed the opposing view to protective legislation, with Section 4 focusing on the printing industry as an example and Section 5 focusing on entertainment industries. Therefore, while content might have caused the time effects, I am more inclined to believe that upon turning the page and
realizing that their task was only half complete, the listeners became somewhat antsy or indifferent. In their boredom or impatience, they found the task tiresome and consequently rated Sections 4 and 5 as harder to understand. The undergraduates lacked the focus to complete the task without getting distracted or letting their mind wander. While I agree that both of these are hypotheses are weak at explaining the ordering effects, they are areas to consider.

The second research question investigated the difference between the ratings for comprehensibility and the ratings for fluency. The results showed a strong correlation between the two with four of the five Spearman coefficients over 0.7. Traditionally, research has separated these two constructs, with most researchers focusing either on comprehensibility or fluency. Comprehensibility is a measure of the amount of work it takes to understand an utterance, and fluency is defined as the amount of automaticity in speech as measured by rate of speed, articulation rate, and other temporal variables. While researchers may separate these two concepts, the results show that in the minds of native speakers these constructs are similar. The differences between comprehensibility and fluency are defined by experts, but to lay people there may be only indistinct differences.

Looking specifically at the different pause treatments, it was hypothesized that correlations between fluency and comprehensibility would be weaker for the pause treatments of frequency and length. This hypothesis was partially supported. Increased frequency, rather than showing a lower correlation, showed the second highest of the treatment correlations at $r = 0.749$ (behind the normal $r = 0.816$). Length, however, did show the lowest correlation between comprehensibility and fluency at $r = 0.612$. Perhaps length of pauses does not influence ratings of comprehensibility, but is rather a part of fluency and the
listener’s intuitive ideas of naturalness and native-like proficiency. Closer inspection of two listeners’ responses supports this conjecture. Two female participants rated the recording with longer length as a ‘2’ for comprehensibility (rather easy to understand) but as a ‘6’ and ‘7’ respectively on fluency (not fluent). Their constructs for comprehensibility and fluency were clearly different, with length violating ideas of fluency but not influencing understanding. In this way we can see that while comprehensibility and fluency have much in common, some temporal aspects may contribute to the perception of one or the other.

5.1 Additional Comprehensibility Items

The third research question examined the strength of Kang’s revised comprehensibility instrument and found that the five averaged items had a strong correlation with the original item at $r = 0.924$. Additionally, this instrument has an internal consistency of 0.94 and the averaging produces an interval item which is more appropriate for statistical testing. While all of these facts recommend the instrument’s future use, we will also consider each individual item in relation to the original descriptor.

5.1.1 Comprehensible/Incomprehensible

It was hypothesized that this item would correlate strongly with “easy/hard to understand.” These two semantic terms seem inseparable, defining each other in the dictionary and appearing in thesaurus groupings. “Understandable” is most directly defined as “comprehensible,” and “comprehensible” mean the “ability to understand.” However, the results do not support this connection. Item 2 at $r = 0.704$ had the lowest correlation with Item 1 among the five items. This perception is further strengthened when looking at the correlations across pause treatments. The item correlations for irregular pauses and frequent pauses were lower than 0.7 ($r = 0.623$ and 0.688 respectively).
One could hypothesize that perhaps the word “comprehensible” had a different meaning in the minds of students. We do not know. I suggest, however, that the instrument design is the primary cause for this low correlation. Three of the five items had the positive item on the right and the negative item on the left. For “comprehensible/incomprehensible” and “clear/unclear” the order is switched. This was done to keep respondents active in their answering and to help prevent respondents from giving the same score to each item. However, the effect was that some students were confused and answered too quickly.

Evidence can be found in the data. One student on a section with more frequent pauses, answered ‘6’ to all of the items, except for item 2. This item was ‘2,’ which would be ‘6’ if inverted. This happened across the data set. For a normal recording, one student answered ‘3’ for three of the items, ‘4’ for the last item, and ‘6’ for the second item. One way to confirm this hypothesis would be to interview participants after the survey to investigate their choices. Another avenue would be to implement two instruments, one following Kang’s (2010) original and one with all of the positive descriptors on the same side. In this way the two instruments could be compared.

If indeed the reversed ordering caused the lower correlations for “comprehensible/incomprehensible,” then how should researchers approach instrument design? It should be noted that the other inverted item “clear/unclear” did not have the same issue. Researchers from advertising to psychology have investigated this issue, and a tenable conclusion is that response order, either positive or negative, does affect participant responses (Weng & Cheng, 2001). For further discussion on this issue consider Belson (1966) and Chan (1991).
5.1.2 Clear/Unclear

This item is interesting because it did not agree with the hypothesis. In fact, instead of showing the weakest correlation to “easy/hard to understand,” this item had the strongest correlation at r = 0.812. This is surprising because while “understanding,” “comprehending,” and “grasping,” are synonymous in meaning, “clear” is the outlier. Of the fifteen definitions for the adjective form of “clear” in *Webster’s College Dictionary*, only one references being easily understood or making meaning clear. Other definitions include bright, light, distinct, serene, calm, obvious, and being free of something (e.g. free from guilt or debt) (Agnes, 2001).

Despite the multiple meanings of “clear,” this descriptor captured the amount of work it took to understand the speaker in a similar fashion to the first item. For example, imagine a non-native speaker is difficult to understand. The native speaker can still understand each word and transcribe them (high intelligibility) but it takes more effort to parse the words of the speaker. The rater would describe them as “unclear” even though they are ultimately intelligible. Therefore, this item is capturing the listening effort, and despite its semantic difference, is getting at the same subjective intuitions as the first item.

5.1.3 Required little/lots of effort and difficult/simple to grasp

Both of these items showed strong and consistent correlations with the original item, with the exception of one pause treatment. “Required little/lots of effort” correlated strongly with all of the treatments except for increased frequency of pauses, and “difficult/simple to grasp” was strong with the exception of length. A closer look at the data was required to see what relationship between the ratings was causing these weaker relationships. For increased frequency, five respondents rated the speaker as ‘2/3’ for “easy/hard to understand” but ‘5/6’
for “required little/lots of effort.” This suggests that the speaker found that they could understand the non-native speaker but it took effort. This item specifically isolated this aspect of comprehensibility that the global descriptor sometimes failed to gather. Rather than showing a weaker correlation to the primary item, I surmise that this type of data is describing the exact type of detail that we seek. Respondents are describing that increased frequency of pauses was easy to understand, but they perceived that it took more effort to process. This insight comes back to the first research question and might explain why increased frequency had the second highest comprehensibility ratings after irregular placement. Overall, these items are not weaker in their relationship to the first item, but rather capture the ideas of comprehensibility in different ways for people who might define the items differently in their minds, and is that not ultimately the objective behind additional items?
CHAPTER 6. CONCLUSION

This study has provided strong evidence for the importance of pause placement and consequently lexical chunks in speech processing. To facilitate the learning of formulaic language, Boers and Lindstromberg (2012) summarized recent research findings and categorized educational approaches into three broad categories: 1) raising awareness; 2) using dictionaries and corpora; and 3) classroom activities. Raising awareness activities included identifying or underlining formulaic language in texts. They also included modifying vocabulary lists to include formulaic language so that students become aware of the importance of multi-word units and not just prioritize single words. The second strategy was teaching dictionary or reference skills to students so that they can learn collocations and chunks on their own. The third category which Boers and Lindstromberg called ‘stimulating retention’ had the largest umbrella. It included rote memorization activities, contrastive-analysis activities such as translation with L1, and proceduralization. Proceduralization implies repetition of a task in order to increase fluency and automatic recall. Proceduralization can also be developed through speaking activities and rote speaking. De Jong and Perfetti (2011) found support for the use of task repetition, and Ding (2007), in interviewing highly successful English learners, found that they attributed their success to repeated recitation. Boers and Lindstromberg identify proceduralization as an area for further research, and this study showing the importance of pause placement supports this proposition. Student need to practice where to breathe and where to break until it becomes automatic. This is one of the key differences between a non-native speaker who is hard to understand and one who is considered fluent and advanced.
Certainly there were limitations to this study. For one, the findings could be strengthened with more responses. The initial experimental design aimed for roughly 100 participants, but the final participant count was 43. This study could be replicated with more responses to make the findings more robust and to counterbalance any outliers or cases in which the participants misread the instrument. Likewise, the content of the lecture might have influenced the ratings. Perhaps participants rated Sections 4 and 5 as hard to understand because of something in the content. Any replication of this study would want to use a different academic lecture to verify that the ratings were not affected by the content of the lecture. A third limitation has already been mentioned, that the speech sample was read speech, not spontaneously produced speech, and this limitation was accepted as part of the methodology. Further limitations are found when looking at the statistical procedures. The ANOVA results would have been stronger with interval rather than ordinal data. Likewise Spearman rank coefficients may overestimate correlations “when there is a high proportion of tied ranks” (Bachman, 2004, p. 90), and this could have affected the results.

Even with all of these limitations, the methodology presented in this study took the work of Hahn (2004) and added another layer. In having all of the students respond to all of the treatments, the researcher was able to factor out the effects of class. This is an important idea in applied linguistics research where so often samples come from predetermined classes. While we might want to assume that these classes are sufficiently randomized because all university undergraduates must take them, they are not, in fact, random. Class formation is affected by the time of day and the teacher, among other factors. We need to keep striving to find experimental research designs, rather than concede that quasi-experimental is good enough.
This study investigated three aspects of pausing: placement, frequency and length. However, future research could consider the difference between filled and unfilled pauses. Pauses can be filled with vocalizations such as *um* or *eh*, and this could affect comprehensibility ratings in a difference why than silent pauses. This study controlled for filled pauses, but the effects of filled and unfilled pauses could be investigated in the future. Also, future research should consider the perceptions of non-native speakers as they form a segment of our classrooms and are part of the EIL (English as an International Language) community. Research can address the question of whether non-native speakers process comprehensibility in the same way as native speakers or in a different way. In this same vein, this study should be replicated with another language to see if these results are consistent across languages or unique to just English.

Another consideration would be to digitize the data collection process as presented in Jaber and Hussein (2011). This has the benefit of randomizing the presentation order. Each participant would receive a different ordering of the 24 unique orders, helping to mitigate ordering effects. Digitizing is also convenient for the researcher because the data is automatically prepared in a spreadsheet. However, would digitizing the survey increase or decrease participation? An argument can be made that computer-based forms are more interactive and engaging, but on the other hand, we are becoming so inundated with digital data that surveys are often deleted as junk mail. It is also harder to guarantee the homogeneity or representativeness of a sample when digital surveys are conducted. Therefore, I leave this question of format for the future.

A final avenue for future research would be to address the limitation in this study of read versus authentic speech. The findings of this study could be confirmed by recording
spontaneous narrative production of non-native speakers, asking native speakers to rate these participants for comprehensibility, and then analyzing the recordings for pause placement to see if there are indeed strong correlations between pause placement and comprehensibility ratings.
APPENDIX A. TRANSCRIPT

Based on MICASE lecture LEL 105SU113
Speech Event type: Lecture-large
Academic Division: Humanities and Arts
Course: History of the American Family
Recorded: October 10, 2000
Word count: 418
http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;cc=micase;view=transcript;id=LEL105SU113

One of the most important issues of the early twentieth century was protective legislation for women. These laws were meant to protect women in the workplace. Women reformers, along with others who were supportive of labor, had the idea that women needed to be protected, if they were going to work in factories. Otherwise their bodies would be damaged for reproduction. And so, protective legislation was a very, very complicated issue.

It’s very interesting that some of the most radical women who were interested in socialism, did not want protective legislation specifically for women. They wanted protective laws for all workers, men and women. But in the end, protective laws were passed, first by the states individually, and then eventually by the federal government.

Now, we believe protective laws did improve the work situation for a lot of women in the factories during this period. They were given breaks. Factory owners were forced to establish women’s bathrooms. Women were prevented from doing certain kinds of labor, night work for example, which people thought mothers shouldn’t be doing. They should be at home taking care of their children. However, in the end, what protective laws accomplished was very, very complicated and mostly negative. It helped women at the immediate moment, when they needed it, but the fact that men weren’t protected also, drove many women workers out of their jobs.
Employers stopped hiring women because it was too costly to make the changes that were needed. For example, there was a large contingent of women, a large percentage of women, working in the printing industry, and most of the printing industry went on at night. Because most of the printing industry’s work was newspapers, women were essentially excluded from the printing industry and lost an opportunity for decent wages and a fair position in industry.

Also, in a number of entertainment industries, bowling for example and amusement parks, women were being hired at fairly decent wages, but they worked at night. Women lost out there, too, and were unable to make gains because in many states those industries were protected. This was also particularly difficult for widows and single mothers, who liked to work at night so that they could be around for their children during the day.

So what we see happening, in the long term with protective laws for women is that it was based on the idea that women are weaker than men, that women need special treatment, instead of equalizing the work experience for all.
APPENDIX B. PAUSE ANALYSIS BETWEEN NATIVE AND NON-NATIVE SPEAKERS

During analysis, each pause placement was given a number. As new pauses appeared in the data, they were given a designation of the nearest number plus ‘a.’ This nomenclature merely made referencing the pauses easier. This chart presents the placement and length of the pauses of the three native speakers and the single non-native speaker. If no pause over 100 milliseconds was produced, the chart is left blank. Some non-native speaker data is missing (X) because the lecture was recorded in chunks, giving the speaker time to break and refresh before continuing. Names have been changed to protect identities.

<table>
<thead>
<tr>
<th>Pause</th>
<th>Location</th>
<th>Kevin</th>
<th>Michelle</th>
<th>Elaine</th>
<th>Harim</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>women. These</td>
<td>0.5</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>S02</td>
<td>workplace. Women</td>
<td>0.3</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>S02a</td>
<td>reformers, along</td>
<td></td>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>S03</td>
<td>labor, had</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>S03a</td>
<td>that women</td>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>S04</td>
<td>protected, if</td>
<td>0.3</td>
<td></td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>S05</td>
<td>factories. Otherwise</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>S05a</td>
<td>Otherwise their</td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>S06</td>
<td>reproduction. And</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>S06a</td>
<td>so, protective</td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>S07</td>
<td>very, very</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S08</td>
<td>issue. It’s</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>X</td>
</tr>
<tr>
<td>S09</td>
<td>women who</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S10</td>
<td>socialism, did</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>S11</td>
<td>women. They</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>S12</td>
<td>workers, men</td>
<td>0.1</td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>S13</td>
<td>women. But</td>
<td>0.3</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>S13a</td>
<td>But in</td>
<td></td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>S14</td>
<td>end, protective</td>
<td></td>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>S15</td>
<td>passed, first</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>S16</td>
<td>individually, and</td>
<td>0.4</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>S17</td>
<td>government. Now,</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>S18a</strong></td>
<td>Now, we</td>
<td></td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td><strong>S19</strong></td>
<td>situation for</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S20</strong></td>
<td>period. They</td>
<td>0.7</td>
<td>0.5</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>S21</strong></td>
<td>breaks. Factory</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>S22</strong></td>
<td>bathrooms. Women</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>X</td>
</tr>
<tr>
<td><strong>S23</strong></td>
<td>labor, night</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>S24</strong></td>
<td>example, which</td>
<td>0.5</td>
<td>0.4</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S25</strong></td>
<td>doing. They</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>S26</strong></td>
<td>children. However,</td>
<td>0.6</td>
<td>0.8</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>S26a</strong></td>
<td>However, in</td>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td><strong>S27</strong></td>
<td>end, what</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td><strong>S28</strong></td>
<td>accomplished was</td>
<td>0.3</td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td><strong>S28a</strong></td>
<td>very, very</td>
<td></td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td><strong>S29</strong></td>
<td>very complicated</td>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td><strong>S30</strong></td>
<td>negative. It</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>S31</strong></td>
<td>moment, when</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>S32</strong></td>
<td>it, but</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>S33</strong></td>
<td>also, drove</td>
<td>0.4</td>
<td>0.2</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S34</strong></td>
<td>jobs. Employers</td>
<td>0.8</td>
<td>0.5</td>
<td>0.6</td>
<td>X</td>
</tr>
<tr>
<td><strong>S34a</strong></td>
<td>women because</td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S35</strong></td>
<td>needed. For</td>
<td>0.4</td>
<td>0.6</td>
<td>1.6</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>S36</strong></td>
<td>example, there</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>S37</strong></td>
<td>women, a</td>
<td>0.1</td>
<td>0.4</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>S37a</strong></td>
<td>women, working</td>
<td>0.2</td>
<td>0.2</td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td><strong>S38</strong></td>
<td>industry, and</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>S39</strong></td>
<td>night. Because</td>
<td>0.6</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>S40</strong></td>
<td>newspapers, women</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>S41</strong></td>
<td>industry and</td>
<td>0.6</td>
<td></td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>S41a</strong></td>
<td>wages and</td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td><strong>S42</strong></td>
<td>industry. Also,</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>X</td>
</tr>
<tr>
<td><strong>S42a</strong></td>
<td>Also, in</td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td><strong>S43</strong></td>
<td>industries, bowling</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>S43a</strong></td>
<td>example and</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S44</strong></td>
<td>parks, women</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>S45</strong></td>
<td>wages, but</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>S45a</strong></td>
<td>but they</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td><strong>S46</strong></td>
<td>night. Women</td>
<td>0.2</td>
<td>0.5</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>S47</td>
<td>too, and</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>S48</td>
<td>gains because</td>
<td>0.1</td>
<td></td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>S49</td>
<td>states those</td>
<td>0.1</td>
<td></td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>S50</td>
<td>protected. This</td>
<td>0.6</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>S50a</td>
<td>mothers, who</td>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>S51</td>
<td>night so</td>
<td></td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S52</td>
<td>day. So</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>S53</td>
<td>happening, in</td>
<td>0.2</td>
<td>0.3</td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>S53a</td>
<td>term with</td>
<td></td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>S54</td>
<td>women is</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>S54a</td>
<td>that is</td>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>S55</td>
<td>men, that</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>S56</td>
<td>treatment, instead</td>
<td>0.6</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Total number</strong></td>
<td><strong>56</strong></td>
<td><strong>51</strong></td>
<td><strong>48</strong></td>
<td><strong>58</strong></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C. VERSION A – NORMAL RECORDING WITH NATIVE-LIKE PAUSES

Pauses rounded to the nearest tenth of a second.

**Section 1 – Baseline (warm-up)**

One of the most important issues of the early twentieth century was protective legislation for women. These laws were meant to protect women in the workplace. Women reformers, along with others who were supportive of labor, had the idea that women needed to be protected, if they were going to work in factories. Otherwise their bodies would be damaged for reproduction. And so, protective legislation was a very, very complicated issue.

**Section 2 (ten pauses)**

It’s very interesting that some of the most radical women who were interested in socialism (.2), did not want protective legislation specifically for women (.7). They wanted protective laws for all workers (.1), men and women (.7). But (.1) in the end (.3), protective laws were passed (.2), first by the states individually (.1), and then eventually by the federal government (.8). Now, we believe protective laws did improve the work situation (.1) for a lot of women in the factories during this period.

**Section 3 (ten pauses)**

They were given breaks (.7). Factory owners were forced to establish women’s bathrooms (.5). Women were prevented from doing certain kinds of labor (.3), night work for example (.1), which people thought mothers shouldn’t be doing (.7). They should be at home taking care of their children (.5). However, in the end (.3), what protective laws accomplished was very, very complicated and mostly negative (.7). It helped women at the
immediate moment (.3), when they needed it (.6), but the fact that men weren’t protected also,
drove many women workers out of their jobs.

Section 4 (nine pauses)

Employers stopped hiring women (.1) because it was too costly to make the changes
that were needed (.7). For example (.4), there was a large contingent of women (.5), a large
percentage of women (.6), working in the printing industry (.7), and most of the printing
industry went on at night (.5). Because most of the printing industry’s work was newspapers
(.4), women were essentially excluded from the printing industry (.3) and lost an opportunity
for decent wages and a fair position in industry.

Section 5 (ten pauses)

Also, in a number of entertainment industries (.3), bowling for example (.1) and
amusement parks (.4), women were being hired at fairly decent wages (.4), but (.1) they
worked at night (.6). Women lost out there, too (.3), and were unable to make gains because
in many states (.2) those industries were protected (.7). This was also particularly difficult
for widows and single mothers (.3), who liked to work at night so that they could be around
for their children during the day.

Section 6 – Second baseline

So what we see happening, in the long term with protective laws for women is that it
was based on the idea that women are weaker than men, that women need special treatment,
instead of equalizing the work experience for all.
APPENDIX D. VERSION B – IRREGULARLY PLACED PAUSES

Pauses rounded to the nearest tenth of a second. Modified pauses in bold. After moving 2/3 of the pauses, pauses of 200 milliseconds were reinserted for naturalness, specifically before paratones in the original recording. These are indicated by a *.

Section 1 – Baseline (warm-up)

One of the most important issues of the early twentieth century was protective legislation for women. These laws were meant to protect women in the workplace. Women reformers, along with others who were supportive of labor, had the idea that women needed to be protected, if they were going to work in factories. Otherwise their bodies would be damaged for reproduction. And so, protective legislation was a very, very complicated issue.

Section 2

It’s very interesting that some of the most radical women who were interested in socialism (.2), did not want protective legislation (.7) specifically for women. (.2*) They wanted protective laws for (.1) all workers, men and women (.7). But in the end, protective laws (.1) were passed (.2), first by the states (.1) individually, and then eventually by the federal government. (.2*) Now, we believe protective laws did (.8) improve the work situation (.1) for a lot of women in the (.3) factories during this period.

Section 3

They were given breaks (.7). Factory owners were forced to establish (.5) women’s bathrooms. Women were prevented from (.3) doing certain kinds of labor, night work for example (.1), which people thought mothers shouldn’t be doing. (.2*) They should be at home taking care of (.5) their children. (.2*) However, in the end (.3), what protective laws accomplished was (.7) very, very complicated and mostly negative. It helped women at the
immediate moment, when they needed it (.6), but the fact that men (.7) weren’t protected also, drove (.3) many women workers out of their jobs.

**Section 4**

Employers stopped hiring women (.1) because it was too costly to make (.7) the changes that were needed. For example, there was a large (.4) contingent of women (.5), a large percentage of women, working in the printing industry, and most of the printing industry went on at (.6) night (.5). Because most of the (.7) printing industry’s work was newspapers, women were essentially excluded from the (.4) printing industry and lost an opportunity for decent wages and a (.3) fair position in industry.

**Section 5**

Also, in a number of entertainment industries (.3), bowling for example and amusement (.1) parks, women were being hired at fairly (.4) decent wages (.4), but they worked at night. Women lost out there (.1), too (.3), and were unable to make gains because in many states those industries were (.2) protected. (.2*) This was also particularly (.6) difficult for widows and single mothers (.3), who liked to work at night so that (.7) they could be around for their children during the day.

**Section 6 – Second baseline**

So what we see happening, in the long term with protective laws for women is that it was based on the idea that women are weaker than men, that women need special treatment, instead of equalizing the work experience for all.
APPENDIX E. VERSION C – INCREASED PAUSE FREQUENCY

Pauses rounded to the nearest tenth of a second. Inserted pauses shown in bold.

Section 1 – Baseline (warm-up)

One of the most important issues of the early twentieth century was protective legislation for women. These laws were meant to protect women in the workplace. Women reformers, along with others who were supportive of labor, had the idea that women needed to be protected, if they were going to work in factories. Otherwise their bodies would be damaged for reproduction. And so, protective legislation was a very, very complicated issue.

Section 2

It’s very interesting that some of the most radical women (.4) who were interested in socialism (.2), did not want protective legislation (.4) specifically for women (.7). They wanted protective laws (.4) for all workers (.1), men and women (.7). But (.1) in the end (.3), protective laws were passed (.2), first by the states individually (.1), and then (.4) eventually by the federal government (.8). Now (.4), we believe protective laws did improve the work situation (.1) for a lot of women in the factories (.4) during this period.

Section 3

They were given breaks (.7). Factory owners were forced to establish women’s bathrooms (.5). Women were prevented from doing certain kinds of labor (.3), night work for example (.1), which people thought (.4) mothers shouldn’t be doing (.7). They should be at home (.4) taking care of their children (.5). However (.4), in the end (.3), what protective laws accomplished (.4) was very, very complicated (.4) and mostly negative (.7). It helped women at the immediate moment (.3), when they needed it (.6), but the fact that men weren’t protected also, drove many women workers (.4) out of their jobs.
Section 4

Employers stopped hiring women (.1) because it was too costly (.4) to make the changes (.4) that were needed (.7). For example (.4), there was a large contingent of women (.5), a large percentage of women (.6), working in the printing industry (.7), and most of the printing industry (.4) went on at night (.5). Because most of the printing industry’s work was newspapers (.4), women were essentially excluded (.4) from the printing industry (.3) and lost an opportunity for decent wages (.4) and a fair position (.4) in industry.

Section 5

Also (.4), in a number of entertainment industries (.3), bowling for example (.1) and amusement parks (.4), women were being hired (.4) at fairly decent wages (.4), but (.1) they worked at night (.6). Women lost out there, too (.3), and were unable to make gains (.4) because in many states (.2) those industries (.4) were protected (.7). This was also particularly difficult for widows and single mothers (.3), who liked to work at night (.4) so that they could be around for their children (.4) during the day.

Section 6 – Second baseline

So what we see happening, in the long term with protective laws for women is that it was based on the idea that women are weaker than men, that women need special treatment, instead of equalizing the work experience for all.
APPENDIX F. VERSION D – PAUSES WITH LONGER LENGTH

Pauses rounded to the nearest tenth of a second. Modified pauses in bold.

Section 1 – Baseline (warm-up)

One of the most important issues of the early twentieth century was protective legislation for women. These laws were meant to protect women in the workplace. Women reformers, along with others who were supportive of labor, had the idea that women needed to be protected, if they were going to work in factories. Otherwise their bodies would be damaged for reproduction. And so, protective legislation was a very, very complicated issue.

Section 2 (ten pauses)

It’s very interesting that some of the most radical women who were interested in socialism (.2), did not want protective legislation specifically for women (1.4). They wanted protective laws for all workers (.2), men and women (.7). But (.2) in the end (.6), protective laws were passed (.2), first by the states individually (.2), and then eventually by the federal government (1.6). Now, we believe protective laws did improve the work situation (.1) for a lot of women in the factories during this period.

Section 3 (ten pauses)

They were given breaks (.7). Factory owners were forced to establish women’s bathrooms (1.0). Women were prevented from doing certain kinds of labor (.6), night work for example (.1), which people thought mothers shouldn’t be doing (1.4). They should be at home taking care of their children (1.0). However, in the end (.3), what protective laws accomplished was very, very complicated and mostly negative (1.4). It helped women at the immediate moment (.6), when they needed it (.6), but the fact that men weren’t protected also, drove many women workers out of their jobs.
Section 4 (nine pauses)

Employers stopped hiring women (.1) because it was too costly to make the changes that were needed (1.4). For example (.8), there was a large contingent of women (.5), a large percentage of women (1.2), working in the printing industry (1.4), and most of the printing industry went on at night (.5). Because most of the printing industry’s work was newspapers (.8), women were essentially excluded from the printing industry (.6) and lost an opportunity for decent wages and a fair position in industry.

Section 5 (ten pauses)

Also, in a number of entertainment industries (.3), bowling for example (.2) and amusement parks (.8), women were being hired at fairly decent wages (.4), but (.2) they worked at night (1.2). Women lost out there, too (.3), and were unable to make gains because in many states (.4) those industries were protected (1.4). This was also particularly difficult for widows and single mothers (.3), who liked to work at night so that they could be around for their children during the day.

Section 6 – Second baseline

So what we see happening, in the long term with protective laws for women is that it was based on the idea that women are weaker than men, that women need special treatment, instead of equalizing the work experience for all.
Thank you for agreeing to participate in this research survey on listening. Your answers will help us understand the listening process, specifically how sounds and patterns can make listening easier or more difficult.

Fill-out the following information:

1. What is your year in school?
   ___ Freshman
   ___ Sophomore
   ___ Junior
   ___ Senior
   ___ Transfer

2. Gender
   ___ Female
   ___ Male

3. Where are you from?
   ___ I’m from Iowa.
   ___ I’m from the Midwest (Minnesota, Wisconsin, Michigan, Illinois, Missouri, Kansas, Nebraska, North and South Dakota, Indiana, and Ohio).
   ___ I’m from the United States, but from another state besides those above.
   ___ I’m from another country.

4. What is your first language?
   ___ English
   ___ Chinese
   ___ Other (Which language? ___________________)
5. Have you had experience abroad, lived abroad or visited a foreign country?
___ Yes, I’ve been abroad.
___ No, I’ve never left the United States.

6. How many of your TAs and professors on campus have been non-native speakers of English? (Write a number):

___

7. Besides having non-native speaking instructors, have you ever had any experience interacting on a regular basis (more than once a week) with a non-native speaker or speakers of English? (Circle one):

Yes   No

8. If you answered Yes to Question 7, how much experience have you had? (Check one):
___ 1-3 months    ___ 1-2 years
___ 4-6 months   ___ 3-4 years
___ 7-12 months  ___ 5 or more years

9. Do you have any known hearing impairments such as hearing loss, buzzing in your ears, or noise-induced hearing loss? (Circle one):

Yes   No

10. Have you read or heard about the historical development of protective legislation for women before? (Circle one):

Yes   No
Directions: After listening to each section, rate the section on the following scales. Do **not** rate while you are listening. Wait until you are finished listening, and then rate the section. There are no right or wrong answers. Just follow your first instinct.

### Section 1
The section I just listened to…

<table>
<thead>
<tr>
<th>was easy to understand</th>
<th>was hard to understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was incomprehensible</td>
<td>was highly comprehensible</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was unclear</td>
<td>was clear</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>required little effort</td>
<td>required lots of effort</td>
</tr>
<tr>
<td>to understand</td>
<td>to understand</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>made it simple to grasp</td>
<td>made it difficult to</td>
</tr>
<tr>
<td>the meaning</td>
<td>grasp the meaning</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>fluent</td>
<td>not fluent</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
</tbody>
</table>

### Section 2
The section I just listened to…

<table>
<thead>
<tr>
<th>was easy to understand</th>
<th>was hard to understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was incomprehensible</td>
<td>was highly comprehensible</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was unclear</td>
<td>was clear</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>required little effort</td>
<td>required lots of effort</td>
</tr>
<tr>
<td>to understand</td>
<td>to understand</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>made it simple to grasp</td>
<td>made it difficult to</td>
</tr>
<tr>
<td>the meaning</td>
<td>grasp the meaning</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>fluent</td>
<td>not fluent</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
</tbody>
</table>

### Section 3
The section I just listened to…

<table>
<thead>
<tr>
<th>was easy to understand</th>
<th>was hard to understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was incomprehensible</td>
<td>was highly comprehensible</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>was unclear</td>
<td>was clear</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>required little effort</td>
<td>required lots of effort</td>
</tr>
<tr>
<td>to understand</td>
<td>to understand</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>made it simple to grasp</td>
<td>made it difficult to</td>
</tr>
<tr>
<td>the meaning</td>
<td>grasp the meaning</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
<tr>
<td>fluent</td>
<td>not fluent</td>
</tr>
<tr>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
<td>___ / ___ / ___ / ___ / ___ / ___ / ___</td>
</tr>
</tbody>
</table>
Section 4
The section I just listened to…
was easy to understand ___ / ___ / ___ / ___ / ___ / ___ / ___ was hard to understand
was incomprehensible ___ / ___ / ___ / ___ / ___ / ___ / ___ was highly comprehensible
was unclear ___ / ___ / ___ / ___ / ___ / ___ / ___ was clear
required little effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
required lots of effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
made it simple to grasp ___ / ___ / ___ / ___ / ___ / ___ / ___ made it difficult to grasp the meaning
made it difficult to ___ / ___ / ___ / ___ / ___ / ___ / ___ understand
fluent ___ / ___ / ___ / ___ / ___ / ___ / ___ not fluent

Section 5
The section I just listened to…
was easy to understand ___ / ___ / ___ / ___ / ___ / ___ / ___ was hard to understand
was incomprehensible ___ / ___ / ___ / ___ / ___ / ___ / ___ was highly comprehensible
was unclear ___ / ___ / ___ / ___ / ___ / ___ / ___ was clear
required little effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
required lots of effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
made it simple to grasp ___ / ___ / ___ / ___ / ___ / ___ / ___ made it difficult to grasp the meaning
made it difficult to ___ / ___ / ___ / ___ / ___ / ___ / ___ understand
fluent ___ / ___ / ___ / ___ / ___ / ___ / ___ not fluent

Section 6
The section I just listened to…
was easy to understand ___ / ___ / ___ / ___ / ___ / ___ / ___ was hard to understand
was incomprehensible ___ / ___ / ___ / ___ / ___ / ___ / ___ was highly comprehensible
was unclear ___ / ___ / ___ / ___ / ___ / ___ / ___ was clear
required little effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
required lots of effort ___ / ___ / ___ / ___ / ___ / ___ / ___ to understand
made it simple to grasp ___ / ___ / ___ / ___ / ___ / ___ / ___ made it difficult to grasp the meaning
made it difficult to ___ / ___ / ___ / ___ / ___ / ___ / ___ understand
fluent ___ / ___ / ___ / ___ / ___ / ___ / ___ not fluent

Thank you so much for your participation! I hope to report my findings to you later this semester.
REFERENCES


ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support of numerous people. First, I would like to recognize my committee. John Levis was always patient and helped me to see the errors in my work through encouraging questions. His suggestions were always sound, and his enthusiasm for research inspires. Gary Phye helped with research design and his meticulous attention to detail has certainly influenced this project. Volker Hegelheimer has been my source of support this semester, and to all three committee members I extend my thanks.

There were also numerous other members of the English department who worked on or advocated this project. They include: Sam Ascoli, Tom Ballard, Dr. Barb Blakely, Corrina Carter, Chloe Clark, Autumn Dinkelman, Brenna Dixon, Clariztbel Flores, Victoria Harthoorn, Ellen Heddendorf, Hong Ma, Meg McCulloch, Dr. Gary Ockey, Mandy Qian, Dr. Geoff Sauer, Karina Silva, Sinem Sonaat, Kirk Wilkins, and Hyejin Yang. There are of course others who have provided support over my two years here at Iowa State University, and while I cannot name them all, I am grateful for the supportive environment engendered by the English department.

Finally, I want to thank my family. My parents have always believed in me, and they supported the pursuit of this degree because they wanted me to be happy and secure in the future. My parents have always supported education; they made this degree possible. For that, I can never thank them enough. The last thanks goes to my husband. He left his home and learned a whole new language just so that I could pursue this degree. He has supported me on all of the long evenings, on the night when I fell asleep reading research articles in my
chair, and on the days when I wanted to give up. He has given me courage with his homemade cookies, his dancing and his laughter. Thank you.