EXAMINING NOISE-INDUCED DIFFERENCES IN THE SYNTACTIC COMPLEXITY
OF SPOKEN LANGUAGE

A Thesis in
Psychology
by
Catherine T. Pham

© 2021 Catherine T. Pham

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2021
The thesis of Catherine T. Pham was reviewed and approved by the following:

Elisabeth A. Karuza
Assistant Professor of Psychology
Thesis Adviser

Janet G. van Hell
Professor of Psychology and Linguistics

Navin Viswanathan
Associate Professor of Communication Sciences and Disorders

Kristin A. Buss
Professor of Psychology and Human Development and Family Studies
Head of the Department of Psychology
ABSTRACT

Speaking in the presence of background noise is a common occurrence in everyday life, yet very few studies have investigated how noise affects speech production beyond the acoustic level. Speaking under noisy conditions is particularly challenging because speakers must not only contend with distracting auditory input but also the knowledge that their listener must now engage in more effortful processing due to transmission degradation of the speech signal. The studies that have examined noise-induced differences in higher-level, non-acoustic properties of speech only presented the noise to the speaker, not the listener. Thus, these studies were unable to investigate the possibility that speakers alter their production to facilitate comprehension for their listener. The present study examined whether (1) background noise affects the syntactic complexity of speech production; and (2) cognitive control predicts noise-induced differences in complexity. Participants completed a picture description task, during which both the speaker and the listener were exposed to the background noise. Overall, speakers reduced the number of clauses, words, and unfilled pauses that they produced in noise, which was correlated with cognitive control. Individuals with weaker cognitive control produced fewer clauses, words, and unfilled pauses in noise. We consider these differences to be speaker-oriented modifications implemented to simplify speech for the speaker’s own benefit. Speakers also reduced the number of filled pauses and mazes in noise, which did not correlate with cognitive control. We consider these differences to be listener-oriented modifications implemented to simplify speech for the listener’s benefit. Thus, we find evidence suggesting that speakers alter the complexity of their speech to alleviate cognitive burden on themselves as well as to facilitate comprehension for their listener.
LIST OF FIGURES

Figure 1. Experimental images for the picture description task .......................................................... 9

Figure 2. Visualizations of each trial types in each cognitive control task ........................................ 13

Figure 3. Two-dimensional scatterplots depicting differences between the noise and silence conditions for each complexity measure .......................................................................................... 21

Figure 4. Bar graphs comparing (A) RTs and (B) percent error for each trial type in the AX-CPT, (B) RTs for the Flanker task, and (C) RTs for the Stroop task ......................................................... 24

Figure 5. Scatterplots depicting correlations between Stroop RT interference and noise-induced differences in (A) number of clauses, (B) words, (C) unfilled pauses, (D) filled pauses, and (E) mazes ......................................................................................................................... 26
LIST OF TABLES

Table 1. Coefficients and corresponding t-values and p-values for the main effect of Condition in a linear mixed effects model examining the effect of Condition and Run on eleven measures of complexity ................................................................. 20
ACKNOWLEDGMENTS

This thesis could not have been completed without the generous time and support of many people. First and foremost, I would like to thank my advisor, Dr. Elisabeth Karuza, for her encouragement, guidance, and unwavering support. I also thank my committee members, Dr. Janet van Hell and Dr. Navin Viswanathan, for their insightful feedback throughout the completion of this thesis. I extend special thanks to Kendra Lange—we are so lucky to have had an incredibly artistic lab manager who was able to create the stimuli for the picture description task. Lastly, I am immensely grateful to Carly Danielson, Brittany Hornberger, and Paige Lemieux for the many hours they spent transcribing the speech production data.
# TABLE OF CONTENTS

**LIST OF FIGURES** .................................................................................................................. iv

**LIST OF TABLES** .................................................................................................................. v

**ACKNOWLEDGMENTS** .......................................................................................................... vi

**Chapter 1. Introduction** ......................................................................................................... 1
  The Effect of Noise on the Acoustic Properties of Speech Production ................................. 2
  The Effect of Noise on the Comprehension of Syntactically Complex Structures ................. 2
  The Effect of Noise on Higher-Level Properties of Speech Production ................................. 3
  The Present Study .................................................................................................................. 5

**Chapter 2. Methods** ............................................................................................................ 8
  Participants ............................................................................................................................... 8
  Stimuli ..................................................................................................................................... 8
    Picture Description Task ........................................................................................................ 8
    Cognitive Control Tasks ........................................................................................................ 9
  Procedure ................................................................................................................................ 10
  Transcription of Speech Production Data ............................................................................. 13
  Measures of Complexity ........................................................................................................ 15
    Ratio-Based Measures .......................................................................................................... 15
    Measures of Quantity ............................................................................................................ 15
    Measures of Disfluencies and Errors ................................................................................... 15

**Chapter 3. Results** ................................................................................................................ 17
  The Effect of Background Noise on Speech Production ......................................................... 17
    Acoustic Analyses ................................................................................................................ 17
    Complexity Analyses .......................................................................................................... 18
  The Role of Cognitive Control in Speech Production in Noise .............................................. 22
    AX-CPT ............................................................................................................................... 22
    Flanker Task ......................................................................................................................... 24
    Number Stroop Task ........................................................................................................... 25

**Chapter 4. Discussion** .......................................................................................................... 28
  Speakers Alter Their Speech Production in Noise ................................................................. 28
  Cognitive Control Modulates Speech Production in Noise ................................................... 30
Points of Convergence and Divergence with Previous Studies .................................................. 32
Qualitative Assessments of Syntactic Complexity ..................................................................... 33
Conclusion ................................................................................................................................. 35

REFERENCES.................................................................................................................................. 37

APPENDIX A. Full Summary of Statistical Analyses ................................................................. 45

APPENDIX B. Transcription Sample.......................................................................................... 48
Chapter 1

Introduction

Spoken language is typically produced in noisy environments. Consider the types of noise a person may encounter in their day-to-day life: a quiet library (30 dB), the humming of a refrigerator (40-50 dB), a classroom full of children (60-90 dB), busy traffic (70-85 dB), or a bustling restaurant (80-90 dB) (Erickson & Newman, 2017). Rarely do we experience true silence, yet psycholinguistic research has largely focused on examining speech production under optimal, relatively silent conditions. The studies that have investigated the effect of noise on speech production have primarily done so at the acoustic level, centering on speakers’ tendencies to increase the loudness of their speech in noisy environments (Lombard, 1911; see also Castellanos et al., 1996; Junqua, 1993; Summers et al., 1988). Although it is clear that speakers alter the acoustic properties of their speech in response to noise, whether they alter the syntactic complexity of their speech remains an open question.

The challenges speakers face in noisy environments are twofold: speakers must contend with (1) the distracting nature of noise coupled with (2) knowledge of the transmission degradation of the speech signal, which results in their listener engaging in more effortful processing. Previous research indicates that sentence comprehension is more difficult in the presence of background noise, especially as structures become more syntactically complex (e.g., Carroll & Ruigendijk, 2013; Wendt et al., 2016). Recent work by Scontras and colleagues (2015) also suggests that complex sentence structures are more difficult to produce than simpler structures. In light of these challenges, we predict that speakers will reduce the complexity of their speech under noisy conditions. To our knowledge, few studies have directly investigated this, despite the pervasiveness of noise in everyday life (but see Harmon et al., 2021; Kemper et al., 2003). The present study compared speech produced in noise versus (relative) silence during
a picture description task to investigate how noise affects syntactic complexity. The present study also examines whether cognitive control modulates speech production in noise.

**The Effect of Noise on the Acoustic Properties of Speech Production**

It is well attested that speakers alter the acoustic properties of their speech in the presence of noise (e.g., Castellanos et al., 1996; Junqua, 1993; Lombard, 1911; Pisoni et al., 1985; Summers et al., 1988). The term Lombard effect refers to an automatic increase in vocal effort when speaking in noise, resulting in perceptually louder speech. Noise-induced acoustic modifications include changes in amplitude, formant frequencies, intensity, and vowel duration (for a review, see Junqua, 1996). Although the Lombard effect is involuntary, the type of background noise modulates the degree to which speakers alter certain properties of their speech (e.g., Garnier et al., 2006; Junqua, 1994; Lu & Cooke, 2008). For example, speakers exhibit a greater increase in vowel duration when speaking in multi-talker babble than white noise (Junqua, 1994). Lombard effects are also modulated by the presence of a communicative task. Garnier et al. (2010) reported that noise-induce modifications were amplified when participants completed a task with a speech partner compared to when did so alone, suggesting that, to some extent, speakers produce Lombard speech to facilitate communication.

**The Effect of Noise on the Comprehension of Syntactically Complex Structures**

Generally, the ways in which noise affects syntactic complexity has been under-studied within the realm of speech production, but studies have examined how noise affects the processing of complex structures (e.g., Carroll & Ruigendijk, 2013; Obleser et al., 2011; Wendt et al., 2016). For example, Carroll and Ruigendijk (2013) investigated the effect of (speech-shaped) noise on the comprehension of German sentences of varying complexity (manipulating ambiguity, canonicity, and embedding of their stimuli). Participants completed a word
monitoring task; each trial began with a target word visually presented on the screen, followed by auditorily presented sentence stimuli. Participants were instructed to push a button as soon as they heard the target word. Crucially, the processing of more complex, non-canonical unambiguous object-verb-subject (OVS) and ambiguous OVS structures were more affected by noise than less complex, canonical subject-verb-object structures. For comparisons of more complex object relative clauses and simpler subject relative clauses, processing difficulties were also greater in noise than silence, suggesting that cognitive demands associated with processing syntactically complex structures are exacerbated in noise.

The Effect of Noise on Higher-Level Properties of Speech Production

Importantly, the handful of studies that have studied the effect of noise on speech production beyond the acoustic level have overlooked the role of the listener, an important consideration given that the Lombard effect is more pronounced in communicative contexts (e.g., Garnier et al., 2010). The studies presented below have solely examined contexts in which the speaker had to contend with the background noise, but the listener did not. Thus, these studies were not poised to reveal noise-induced changes related to the speaker’s knowledge of the transmission degradation of the speech signal. Kemper and colleagues (2003) examined how the presence of background noise affected fluency, syntactic complexity, and propositional content (i.e., amount of information conveyed) in speech production in younger and older adults. Participants answered primarily autobiographical elicitation questions while ignoring speech (a recording of an individual reading semantically anomalous sentences) and while ignoring noise (a recording of ambient cafeteria noise). Participants exhibited decreased speech rates while speaking in noise, decreased syntactic complexity, and decreased propositional content. Differences were greater when ignoring speech than ignoring ambient cafeteria noise.
Harmon et al. (2021) also investigated the effect of noise on higher-level properties of speech production, specifically probing the effect of five different types of background noise (a two-person debate, dialogue from a movie, contemporary music with English lyrics, classical music with Latin vocals, and pink noise). Participants answered prompts that they selected for themselves from a list of available options. Notably, participants exhibited increased disfluencies across all noise conditions (with the exception of pink noise) and fewer unfilled pauses across all noise conditions (with the exception of classical noise). Participants also produced more ungrammatical utterances in noise (i.e., more lexical-phonological, morpho-syntactic, and macro-linguistic errors such as incomplete or tangential utterances), but the effect of noise on the production of ungrammatical utterances varied across the different noise conditions.

Taken together, these two studies (Harmon et al., 2021; Kemper et al., 2003) provide evidence that noise does affect non-acoustic properties of speech production. However, in both studies, background noise was presented to participants through headphones, so the experimenter (the listener) was not privy to the noise. Consequently, these studies were unable to examine the effect of the listener on speech production in noise. Although the speaker may have been distracted by the noise, it is unlikely they altered their speech for the benefit of their listener. Thus, as described in further detail below, a benefit of the present study’s method of data collection through Zoom is that both the participant and the experimenter could be exposed to the background noise without sacrificing the quality of the audio recording, allowing us ask whether participants might also alter the complexity of their speech for the benefit of their listener. Furthermore, neither of the previously mentioned studies directly examined how individual differences in cognitive control may modulate the effect of noise on speech production. Given that speaking in noise entails ignoring distracting auditory input, speakers
with greater cognitive control may be better able to contend with the distracting nature of noise and exhibit fewer differences between speech produced in noise versus silence. Thus, we included measures of cognitive control as a means of exploring potential forces behind noise-induced differences in complexity.

The Present Study

The present study investigated whether speakers alter the syntactic complexity of their speech in the presence of background noise during a picture description task. Our measures of syntactic complexity were divided into three categories: ratio-based measures, measures of quantity, and measures of disfluencies and errors. Our ratio-based measures (mean length of T-unit (MLTU) and clausal density (CD)) are arguably the most commonly utilized metrics of syntactic complexity (Jagaiah et al., 2020). Both MLTU and CD are commonly utilized in the fields of communication sciences and disorders (e.g., Nippold, 2014) and applied second language acquisition (e.g., Lu, 2011), with higher MLTU and CD considered to be reflective of greater syntactic complexity. Measures of quantity (number of T-units, clauses, and words) were also examined. Although these are not traditionally considered measures of complexity per se, they capture whether people generally speak less in the noise condition, which we would expect given the presumed cognitive burden of speaking in noise. Syntactic complexity can also be operationalized in terms of the effort required to produce or comprehend a particular structure. Disfluencies most often appear where greater demand on utterance planning emerges, and the rate of disfluencies increases as utterance planning difficulty increases (Bortfeld et al., 2001). Complex structures have been argued to require more cognitive resources to plan (MacDonald, 2013) and produce (Scontras et al., 2015). Thus, increased rates of disfluencies could be due to the production of complex syntactic structures or difficulties in utterance planning in noisy
environments. We predicted that speech produced in the context of noise will be less syntactically complex than speech produced in silence, reflected by lower MLTU and CD and reduced quantity of speech (i.e., fewer T-units, clauses, and words). Because speaking in noise is presumably cognitively burdensome, we predicted more disfluencies and errors in noise relative to silence.

The present study also explored the extent to which noise-induced changes in syntactic complexity were modulated by individual differences in cognitive control. To produce fluent and grammatically correct speech, speakers must suppress task-irrelevant stimuli and engage in a monitoring process that has been argued to involve some degree of cognitive control (Piai et al., 2013); this demand is likely to be even greater against background noise. Three tasks were selected to assess different aspects of cognitive control: an AX-Continuous Performance Task (AX-CPT; Braver et al., 2001), a Flanker task (Eriksen & Eriksen, 1974), and a counting Stroop task (Bush et al., 2006). The AX-CPT allows for the examination of proactive control and reactive control. Proactive control involves anticipating and preventing interference before it can occur whereas reactive control involves detecting and resolving interference after it occurs (Braver, 2012). Within the context of speaking in noise, speakers may implement proactive control, preparing to alter their speech in anticipation of noise-induced disruptions in utterance planning, or they may implement reactive control, altering their speech only when noise disrupts their utterance planning process. The Flanker and counting Stroop tasks involve overcoming interference effects and reflect the suppression of task-irrelevant stimuli (i.e., noise). The counting Stroop task is more linguistic in nature and involves an arguably more prepotent incorrect response compared to the Flanker (Nee et al., 2007). Thus, by examining performance on these two tasks, we intended to compare linguistic and non-linguistic cognitive control. We
predicted individuals with greater cognitive control would be less susceptible to the additional
cognitive burdens of speaking in noise and would exhibit fewer differences in complexity for
speech produced in noise relative to silence.
Chapter 2

Methods

Participants

Sixty-two participants recruited from the Pennsylvania State University’s Psychology SONA subject pool participated in this study (56 females, 6 males; Age: $M = 18.90$ years, $SD = 1.18$ years). In exchange for their participation, participants received one research credit. All participants were monolingual native English speakers with self-reported normal or corrected-to-normal vision and hearing. Overall, data from five participants were excluded: three due to the presence of extraneous background noise noticed by the experimenter, one due to the completion of the study while wearing a face mask, and one due to audio recording problems. Fifty-seven participants were included in the speech production analyses.

A priori power analysis indicated that a sample size of thirty-four participants would be sufficient to detect a medium effect ($d = 0.5$) for a within-group comparison of speech produced in noise compared to silence (using a Type I error rate of 0.05 and a Type II of 0.2). Additionally, prior work by Vuong & Martin (2014) found a significant correlation ($r = 0.37$) between cognitive control and comprehension of garden path sentences, which are typically considered to be complex sentences. A priori power analysis indicated that a sample size of fifty-five participants would be sufficient for similar planned correlational analyses in the present study. Thus, our final sample size of fifty-seven participants was adequate for our analyses.

Stimuli

Picture Description Task

The picture description stimuli were comprised of one practice image and four experimental images. The experimental images were created to emulate the classic “Cookie Theft” image from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1972).
Each image depicts some overarching scene with many supporting details. In an adaptation of the criteria used by Catricalà and colleagues (2017), each of the experimental images depicts the following: eight subjects (all animate nouns), ten actions (at least two transitive and two ditransitive), and thirteen unique objects (see Figure 1).

![Experimental images for the picture description task.](image)

**Cognitive Control Tasks**

For the AX-CPT, each trial consisted of a series of five letters: a red cue letter, followed by three black distractor letters, and a red probe letter. There were four trial types: target AX trials (where A appeared as the cue and X appeared as the probe) and non-target AY trials (where A appeared as the cue and a non-X letter of the alphabet appeared as the probe), BX trials
(where a non-A letter of the alphabet appeared as the cue and X appeared as the probe), and BY trials (where neither the cue nor the probe were A or X). The letter A never appeared in the probe position, and the letter X never appeared in the cue position. Furthermore, the letters K and Y never appeared during the letter sequence due to their perceptual similarity to the letter X.

The Flanker task consisted of congruent and incongruent trials. Each trial contained a row of five angle brackets. For the congruent trials, the center bracket and the four flanking brackets faced the same direction (e.g., < < < < < ). For the incongruent trials, the center bracket and the four flanking brackets faced opposite directions (e.g., < < > < < ). Thus, the incongruent trials required suppression of the flanking brackets.

The counting Stroop task consisted of neutral and incongruent trials. For the neutral trials, the words on the screen were all monosyllabic animal words (e.g., cat cat). For the incongruent trials, the words on the screen were monosyllabic number words (e.g., one one one). Thus, incongruent trails required the suppression of the semantic content of the number words.

**Procedure**

After providing verbal consent, participants completed the picture description task. We utilized a picture description task over other common speech elicitation techniques (e.g., expository discourse tasks) because it allows for more control over the semantic content of participants’ productions. Controlling the semantic content allowed for a greater focus on the syntax of participants’ productions. The picture description task was administered over Zoom with the experimenter present. Participants were explicitly asked to wear headphones for the picture description task. During this task, participants orally described the set of images in as much detail as possible. The images were described against (relative) silence or noise, alternating the two conditions for each image. During the noise condition, sixteen-talker babble (Harris,
2018) consisting of eight male talkers and eight female talkers was played over participants’ headphones. This babble was unintelligible and devoid of syntactic and semantic information but maintained speech-like acoustic properties, making it more cognitively demanding to listen to than other types of noise (e.g., white noise) (e.g., Danhauer & Leppler, 1979). The order of the experimental images and the noise and silence conditions was counterbalanced across participants.

Before starting the picture description task, the experimenter worked with the participant to test the loudness of the background noise. The experimenter played the background noise audio file on their computer and shared the background noise with the participant through Zoom. While wearing headphones, participants were asked to adjust the volume of their own computer until they felt that the loudness of the noise had reached a “loud party level,” meaning that the noise was loud enough to be distracting and difficult (but not impossible) to converse in. Crucially, the experimenter made it clear to the participant that they were also able to hear the background noise by increasing vocal intensity while speaking to the participant during the background noise adjustment process. Once the volume level was adjusted, participants were explicitly told not to change the volume of their computer for the duration of the study.

Participants were instructed to describe each image for at least two minutes and thirty seconds, but no more than five minutes. A small timer was present in the lower right-hand corner of the screen so that participants could monitor how long they had been describing each image. During the picture description task, participants were informed that the experimenter would be taking notes by hand as the participant described each image to make the task feel more communicative and encourage the participant to consider their addressee’s (i.e., the experimenter’s) communicative needs during the task. Crucially, participants were aware that the
experimenter could hear the background noise. Participants’ speech was recorded using the audio record function in Zoom.

Following the picture description task, participants completed a battery of cognitive control tasks: the Flanker task, the counting Stroop task, and the AX-CPT. The cognitive control tasks were administered online via Pavlovia, an online platform for hosting experiments built in PsychoPy (Peirce et al., 2019). The order of the cognitive control tasks was counterbalanced across participants. Following the protocol outlined in Karuza et al. (2016), the Flanker task consisted of an equal number (n = 188) of congruent and incongruent trials. Participants were instructed to press either the right or left arrow keys to indicate which direction the center bracket was facing (see Figure 2A). Each trial was displayed for 800 ms followed by a fixation cross with an ISI interval drawn from a uniform distribution (500–1250 ms). Congruent and incongruent trials were randomly interleaved, and reaction times and accuracy were recorded.

The counting Stroop task (Bush et al., 2006) consisted of 8 blocks of 20 trials with alternating neutral and incongruent blocks. Participants were instructed to count the total number of words presented on the screen, ranging from 1 to 4, and push the corresponding number key on their keyboard (see Figure 2B). Each trial was displayed for 1500 ms with an ISI interval drawn from a uniform distribution (500–1250 ms). Reaction times and accuracy were recorded.

The AX-CPT (Braver et al., 2001) consisted of 80 trials. Each trial consisted of a series of five letters presented one at a time on the center of the screen: a red cue letter, followed by three black distractor letters, and a red probe letter. Following all cue and distractor letters, participants were instructed to press the right arrow key. Following the probe letter, participants were asked to change their response and press the left arrow key if the probe was an X that was preceded by an A cue (see Figure 2C). Otherwise, participants were instructed to press the right
arrow key. 70% of all trials consisted of AX trials (n = 56), and 30% consisted of an equal number of trials (n = 8) for the remaining three trial types (AY, BX, and BY). Reaction times and accuracy were recorded.

Figure 2. Visualizations of each trial types in each cognitive control task. During the Flanker task, participants responded to the direction of the middle arrow as shown in (A). During the Stroop task, participants responded to the number of words on the screen, as shown in (B). During the AX-CPT, participants pressed the right arrow key in response to all cue and distractor letters. Crucially, in response to the probe letter, participants only pressed the left arrow key if the red cue-probe pair corresponded to an AX pair, as shown in (C).

Transcription of Speech Production Data

Production data were transcribed verbatim by the author or a research assistant. Because it is notoriously difficult to identify sentence boundaries in spoken language, the transcriptions were segmented into minimal terminable units (T-units), which consist of a main clause and any
subordinate clauses or nonclausal structures attached to it (Hunt, 1970). For utterances that contained coordinate main clauses, if a subject was explicitly produced in the first main clause but not in the second clause the utterance was considered one T-unit (e.g., *the woman is petting a seal and is holding a drink in her hand*). However, if a subject was explicitly produced in both clauses, the utterance was segmented into two T-units (e.g., *the woman is petting a seal / and she is holding a drink in her hand*). Utterances that lacked an explicit subject and/or a verb were considered fragments and placed in parentheses and excluded from analysis. However, if the subject of the utterance could be inferred based on previous context, that utterance was considered a T-unit (e.g., *the older woman could be her mother or grandmother / (she) has shoulder_length hair*). Clauses were coded in accordance to the guidelines from Nippold and colleagues (2014), identifying main, adverbial, relative, nominal, infinitive, participial, and gerund clauses. Pairs of words were treated as compound words if they were included as lexical entries in Merriam-Webster’s Dictionary, in which case these words would be linked using an underscore (e.g., *beach ball* transcribed as *beach_ball*). While an imperfect method (e.g., *beach ball* is listed as a lexical entry but *soccer ball* is not), this approach ensured that semantically linked word pairs would be transcribed consistently across participants. Proper nouns were also treated as one word and linked with an underscore (e.g., *Loch Ness Monster* transcribed as *Loch_Ness_Monster*). Disfluencies (i.e., (un)filled pauses, mazes) were placed in parentheses. Discourse markers (e.g., *okay, yeah*) were also placed in parentheses and excluded from analysis. Grammatical errors were annotated (see the discussion of errors under the Measures of Disfluencies and Errors section below). All transcription files were doubled checked by at least one other member of the research team.
Measures of Complexity

Below we briefly describe eleven complexity metrics that were extracted from each participant’s speech production data. We have divided these measures into three categories: ratio-based measures, measures of quantity, and measures of disfluencies and errors.

Ratio-Based Measures

Our ratio-based measures consisted of MLTU and CD. MLTU was calculated by dividing the total number of words by the total number T-units in a language sample. CD was calculated by dividing the total number of clauses (main and subordinate) by the total number of T-units (Hunt, 1965). Disfluencies were removed before calculating MLTU and CD.

Measures of Quantity

The next set of metrics that we examined are the average number of T-units, clauses, and words produced by participants in each condition, after the removal of all filled pauses and mazes.

Measures of Disfluencies and Errors

We divided the category of disfluencies into two subcategories: pauses and mazes. For our analyses, we make the distinction between unfilled (silent) pauses and filled pauses (um, uh, hm, ah, er) because the two been argued serve different functions in speech production. Under some accounts, unfilled pauses are considered a proxy for a speaker’s difficulty in utterance planning (e.g., Goldman-Eisler, 1968), whereas filled pauses are thought to be produced for the listener’s benefit (e.g., Clark & Fox Tree, 2002). Thus, we calculated the average number of unfilled and filled pauses separately. Additionally, we also calculated total unfilled pause duration to account for the possibility that participants may differ in how long they pause in each condition. In accordance with Goldman-Eisler's (1968) standard 250-ms threshold for
distinguishing between pauses associated with articulation (< 250 ms) and hesitations (≥ 250 ms) in speech production, our calculations for number of unfilled pauses and duration of unfilled pauses only included pauses greater than or equal to 250 ms in length. In addition to pauses, our other measure of disfluency was number of mazes, which encompasses abandoned words/utterances, false starts, repetitions, and reformulations. Adjacent mazes were combined, so if multiple mazes appeared in row, they were placed within the same set of parentheses and only counted once.

We also measured the number of errors participants produced. Although participants did produce some phonetic and semantic errors during the picture description task, we only focused on syntactic errors in the present analyses. Syntactic errors included incorrect subject-verb agreement (e.g., there is hills in the background), the production of an unnecessary contraction (e.g., she has a purse with her that’s looks very full), the insertion of an unnecessary word that renders the T-unit ungrammatical (e.g., it looks like mountains in the background which makes me to believe that it’s a lake), or the omission of a single word (e.g., I’m gonna assume it’s (a) pretty windy day). Although we initially did not intend to conduct a separate analysis of number of omissions, we opted to do so after seeing the prevalence of both functional word omissions (e.g., I’m gonna assume it’s (a) pretty windy day) and subject omissions (e.g., the older woman could be her mother or grandmother / (she) has shoulder_length hair) in the data.
Chapter 3
Results

The Effect of Background Noise on Speech Production

**Acoustic Analyses**

Given that data were collected remotely through Zoom, one potential concern is the lack of control the experimenter had over the participants’ noise levels. Thus, because it is well attested that speakers exhibit the Lombard effect when speaking in noisy environments (e.g., Castellanos et al., 1996; Junqua, 1993; Lombard, 1911; Pisoni et al., 1985; Summers et al., 1988), we examined whether our participants exhibited evidence of speaking louder during the noise condition relative to silence condition. Generally, greater intensity in the noise relative to silence condition would serve as strong evidence that participants were complying with task instructions. We used Praat (v. 6.1.41; Boersma & Weenink, 2021) to calculate mean intensity and peak intensity of participants’ speech production. Because participants varied in the number and duration of unfilled (silent) pauses that they produced, we calculated mean intensity with pauses included as well as with pauses with a duration of greater than 250 ms excluded.

Overall, we found that participants exhibited higher mean intensities in the noise condition (without pauses: \(M = 70.80\) dB, \(SD = 1.90\) dB; with pauses: \(M = 69.39\) dB, \(SD = 1.86\) dB) than in the silence condition (without pauses: \(M = 69.89\) dB, \(SD = 1.74\) dB; with pauses: \(M = 68.41\) dB, \(SD = 1.73\) dB). Mean intensities were significantly different between the two conditions (without pauses: \(t(56) = 5.46, p < 0.00001\); with pauses: \(t(56) = 5.23, p < 0.00001\)). Participants also exhibited a higher peak intensity in noise (\(M = 82.10\) dB, \(SD = 1.28\) dB) compared to silence (\(M = 81.69\) dB, \(SD = 1.54\) dB). Peak intensity also significantly differed between the two conditions (\(t(56) = 2.57, p = 0.01\)). In sum, when considering all three types of
intensity measurements (mean intensity with pauses removed, mean intensity with pauses included, and peak intensity), participants generally spoke louder in the noise condition compared to silence.

Because the effect of noise on speech production varies across speakers (Junqua, 1993), we differentiated between participants who were more strongly affected by the background noise (at least at the acoustic level) and those who were not. We calculated the difference in mean intensity (with pauses removed) between the noise and silence condition for each participant. Forty-six participants had positive mean intensity differences (indicating that they spoke louder in the noise condition) and eleven participants had negative mean intensity differences (indicating they spoke louder in the silence condition). Although the results presented below include all fifty-seven participants, we also ran the analyses reported below with the subset of forty-six participants who spoke louder in the noise condition, and the same pattern of results hold.

**Complexity Analyses**

In the present study, our primary comparison of interest was the difference between speech produced in the noise relative to the silence condition. For each of our eleven complexity metrics, we implemented a linear mixed effects model (LMM) using the `lmer()` function in the lme4 package (v. 1.1-26; Bates et al., 2015) in R (v. 4.0.4; R Core Team, 2021). Statistical significance was evaluated using the R lmerTest package (v. 3.1-3; Kuznetsova et al., 2017). Each complexity metric was regressed onto all main effects and interactions of Condition (noise versus silence) and Run (1-4), which was included to account for the possibility that participants’ performance during the picture description task might improve as they described more images. Each model included the fullest random effects structure that allowed the model to converge. For
Model 8 (Filled Pauses), Model 9 (Mazes), and Model 11 (Omissions), the fullest random effects structure corresponded to a random intercept for Participant. For Model 1 (MLTU), Model 2 (CD), Model 3 (T-units), Model 7 (Pause Duration), and Model 10 (Errors), the fullest random effects structure corresponded to a random intercept for Participant and for Item. For Model 4 (Clauses) and Model 5 (Words), the fullest random effects structure corresponded to a random intercept for Participant and Picture ID and a by-participant random slope for Run. For Model 6 (Unfilled Pauses), the fullest random effects structure corresponded to a random intercept for Participant and Picture ID and by-participant random slope for Condition and Run.

Because we were principally interested in the main effect of Condition on our complexity measures, we summarize those results in Table 1. For the full summary of the analyses, see Table A1 in Appendix A. Contrary to our original prediction, we did not observe a significant effect of Condition for our ratio-based measures of complexity. Participants did not produce longer T-units (MLTU: $\beta = 0.10, t = 1.02, p = 0.31$) nor did they produce more clauses per T-unit in silence compared to noise (CD: $\beta = 0.02, t = 1.24, p = 0.22$). For our measures that captured the overall quantity of speech produced, we observed a marginal main effect of Condition for number of T-units ($\beta = 0.59, t = 1.76, p = 0.080$) and a significant main effect of Condition for number of clauses ($\beta = 1.47, t = 3.25, p = 0.001$) and words ($\beta = 10.94, t = 3.43, p < 0.001$). Notably, participants produced more clauses and words in silence compared to noise (Figure 3). Regarding our measures of disfluencies, we observed a main effect of Condition for number of filled pauses ($\beta = 0.79, t = 3.38, p < 0.001$), unfilled pauses ($\beta = 2.59, t = 4.30, p < 0.0001$), and mazes ($\beta = 0.63, t = 2.54, p = 0.01$), indicating that participants produced more filled pauses, unfilled pauses, and mazes when speaking in silence compared to noise (Figure 3). Crucially, the significant effect of Condition for number of filled pauses, unfilled pauses, and mazes was
maintained even after accounting for number of clauses and words (filled pauses: $\beta = 0.66, t = 2.82, p = 0.005$; unfilled pauses: $\beta = 1.80, t = 3.50, p < 0.001$; mazes: $\beta = 0.50, t = 2.00, p = 0.047$). In other words, participants’ tendency to produce more filled pauses, unfilled pauses, and mazes in noise cannot solely be attributed to the fact that they also produced more clauses and words in noise. The effect of pause duration was marginal ($\beta = 0.80, t = 1.84, p = 0.07$).

Participants did not exhibit a significant difference in the number of errors ($\beta = 0.02, t = 0.17, p = 0.87$) nor the number of omissions ($\beta = 0.03, t = 0.28, p = 0.78$) that they produced in noise relative to silence.

Table 1. Coefficients and corresponding t-values and p-values for the main effect of Condition in a linear mixed effects model examining the effect of Condition and Run on eleven measures of complexity. Significant values are bolded. To see the full summary of the results, including the main effect of Run and the interaction between Condition and Run, refer to Table A1 in Appendix A.

<table>
<thead>
<tr>
<th>Complexity Measure</th>
<th>Coefficient</th>
<th>$T$-value</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL 1: Mean Length of T-unit</td>
<td>0.10</td>
<td>1.02</td>
<td>0.31</td>
</tr>
<tr>
<td>MODEL 2: Clausal Density</td>
<td>0.02</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td>MODEL 3: Number of T-units</td>
<td>0.59</td>
<td>1.76</td>
<td>0.08</td>
</tr>
<tr>
<td>MODEL 4: Number of Clauses</td>
<td><strong>1.47</strong></td>
<td><strong>3.25</strong></td>
<td><strong>&lt; 0.001</strong></td>
</tr>
<tr>
<td>MODEL 5: Number of Words</td>
<td><strong>10.94</strong></td>
<td><strong>3.43</strong></td>
<td><strong>&lt; 0.001</strong></td>
</tr>
<tr>
<td>MODEL 6: Number of Unfilled Pauses</td>
<td><strong>2.59</strong></td>
<td><strong>4.30</strong></td>
<td><strong>&lt; 0.0001</strong></td>
</tr>
<tr>
<td>MODEL 7: Total Pause Duration</td>
<td>0.80</td>
<td>1.84</td>
<td>0.07</td>
</tr>
<tr>
<td>MODEL 8: Number of Filled Pauses</td>
<td><strong>0.79</strong></td>
<td><strong>3.38</strong></td>
<td><strong>&lt; 0.001</strong></td>
</tr>
<tr>
<td>MODEL 9: Number of Mazes</td>
<td><strong>0.63</strong></td>
<td><strong>2.54</strong></td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>MODEL 10: Number of Errors</td>
<td>0.02</td>
<td>0.17</td>
<td>0.87</td>
</tr>
<tr>
<td>MODEL 11: Number of Omitted Words</td>
<td>0.03</td>
<td>0.28</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Figure 3. Two-dimensional scatterplots depicting differences between the noise and silence conditions for each complexity measure. Each point represents one participant. The diagonal line represents a zero difference between the noise and silence condition. The greater the distance between a point and the diagonal line, the greater the difference between the noise and silence condition for that individual participant. The red boxes signify complexity measures that exhibited a significant main effect of Condition \((p \leq 0.01)\).
The Role of Cognitive Control in Speech Production in Noise

Because we were also interested in examining the effect of cognitive control on speech production in noise versus silence, we computed correlations between the cognitive control measures and the five complexity measures that significantly differed between the noise and silence conditions (Clauses, Words, Unfilled Pauses, Filled Pauses, and Mazes).

**AX-CPT**

Prior to analysis of the AX-CPT data, we excluded participants who did not meet an overall accuracy threshold of 70%. This approach led to the exclusion of 5 participants ($N = 52$). After the exclusion of these participants, overall performance on the AX-CPT was relatively high (mean accuracy = 90.12%, $SD = 0.06$). Regarding the RT data, we excluded RTs less than 200 ms (5.79% data loss) and, because RTs were not capped during the AX-CPT, we also excluded RTs greater than 1300 ms (1.59% data loss), which was cutoff used by Braver et al. (2001). Analyses were conducted utilizing RTs on correct trials. In light of the right skew of the AX-CPT RT data (skewness = 1.68), RTs were log-transformed before analysis.

During the AX-CPT, participants may have adopted different control strategies while completing the task. Some participants may have relied on proactive goal maintenance, preparing key responses to the probe after encountering the cue. For BX trials, this strategy would be helpful because encountering a B cue would signify that the current trial will be a non-target trial. However, this strategy would not be helpful for AY trials, because participants would have incorrectly prepared for a target probe following the A cue, resulting in longer RTs and more errors. Other participants may have relied on reactive inhibitory control, not preparing key responses to the probe after encountering the cue, resulting in longer RTs and more errors for BX trials. To quantify individual participants’ bias towards one control strategy over the other, we
calculated the Behavioral Shift Index (BSI; Braver et al., 2009) for RTs using the following formula \((AY - BX)/(BX+BY)\). Using z-score transformed RTs and percent error, we calculated a composite BSI score. A positive BSI score indicates reliance on proactive control, whereas a negative BSI score indicates reliance on reactive control.

Focusing on comparisons between our two principal trial types of interest, we found that participants had longer RTs on AY trials \((M = 6.28, SD = 0.23)\) than on BX trials \((M = 5.77, SD = 0.20)\). RTs were significantly different between these two trial types \((t(51) = 15.00, p < 0.00001)\) (Figure 4A). Participants also exhibited a higher percentage of errors for AY trials \((M = 0.23, SD = 0.17)\) than BX trials \((M = 0.05, SD = 0.07)\), which was significantly different between the two trial types \((t(51) = 6.46, p < 0.00001)\) (Figure 4B).

For each of our five complexity measures that exhibited a significant main effect of Condition (Clauses, Words, Unfilled Pauses, Filled Pauses, and Mazes), we examined whether the magnitude of the difference between the noise and silence conditions was correlated with BSI. We did not find evidence of a correlation between these five measures and BSI scores (see Table A2 in Appendix A). Neither a reliance on proactive control nor a reliance on reactive control significantly correlated with participants’ tendency to produce fewer clauses, words, unfilled pauses, filled pauses, and mazes in noise compared to silence. However, recall that our a priori power analysis indicated that a minimum sample size of fifty-five participants is needed to detect a correlation between cognitive control and syntactic complexity. For the AX-CPT, we had a final sample size of fifty-two participants, which is less than the minimum sample size. As such, these analyses were underpowered, which might explain why we did not observe any significant correlations between BSI scores and our measures of complexity.
Figure 4. Bar graphs comparing (A) RTs and (B) percent error for each trial type in the AX-CPT, (B) RTs for the Flanker task, and (C) RTs for the Stroop task. As these represent within-participant measures, error bars are not included.

**Flanker Task**

Prior to analysis of the Flanker data, we excluded participants who did not meet an overall accuracy threshold of 70%. This approach led to the exclusion of 13 participants (total remaining participants for analysis, N = 44). After the exclusion of these participants, overall performance on the AX-CPT was quite high (mean accuracy = 90.12%, SD = 0.05). Regarding the RT data, we excluded RTs less than 200 ms (3.92% data loss). Analyses were conducted utilizing RTs on correct trials.

As expected, participants exhibited the longer RTs for incongruent trials (M = 6.64, SD = 0.12) compared to congruent trials (M = 6.58, SD = 0.12) (Figure 4C). RTs were significantly different between the two trial types (t(43) = 14.61, p < 0.00001). For each of our five complexity measures that exhibited a significant main effect of Condition (Clauses, Words, Unfilled Pauses, Filled Pauses, and Mazes), we examined whether the magnitude of the
difference between the noise and silence conditions was correlated with the RT interference effect (obtained by calculating the difference in RTs between incongruent and congruent trials). We did not find evidence of a correlation between these five measures and RT interference (see Table A2 in Appendix A). RT interference during the Flanker task was not significantly correlated with participants’ tendency to produce fewer clauses, words, unfilled pauses, filled pauses, and mazes in noise compared to silence. However, as was the case with the AX-CPT, our Flanker analyses were underpowered with a final sample size of only forty-four participants, which might explain why we did not find any significant correlations between RT interference and measures of complexity.

**Number Stroop Task**

All participants achieved an overall accuracy score of at least 70%, so we did not exclude any participants from the Stroop analyses (N = 57). Overall performance on the Stroop task was quite high (mean accuracy = 92.87%, SD = 0.04). Regarding the RT data, we excluded RTs less than 200 ms (1.58% data loss). Analyses were conducted utilizing RTs on correct trials. In light of the right skew of the Stroop RT data (skewness = 1.06), RTs were log-transformed before analysis.

As expected, participants exhibited the longer RTs for incongruent trials (M = 6.64, SD = 0.12) compared to neutral trials (M = 6.58, SD = 0.12) (Figure 4D). RTs were significantly different between the two trial types (t(57) = 9.91, p < 0.00001). For each of our five complexity measures that exhibited a significant main effect of Condition (Clauses, Words, Unfilled Pauses, Filled Pauses, and Mazes), we examined whether the magnitude of difference between the noise and silence conditions was correlated with the RT interference effect (obtained by calculating the difference in logRTs between incongruent and neutral trials) on the Stroop task. Stroop RT
interference was not significantly correlated with differences in number of filled pauses ($r = -0.20$, $p = 0.13$) or mazes ($r = 0.02$ $p = 0.86$), but it was significantly correlated with differences in number of clauses ($r = -0.33$, $p = 0.01$), words ($r = -0.29$, $p = 0.03$), and unfilled pauses ($r = -0.32$, $p = 0.02$). Individuals who exhibited larger interference effects, suggestive of weaker cognitive control, produced fewer clauses, words, and unfilled pauses when speaking in noise than silence (Figure 5). Although the correlation analyses were planned beforehand, we note that the significance of these correlations do not hold when subjected to a conservative multiple comparisons correction (i.e., the Bonferroni correction).

Figure 5. Scatterplots depicting correlations between Stroop RT interference and noise-induced differences in (A) number of clauses, (B) words, (C) unfilled pauses, (D) filled pauses, and (E) mazes. The red boxes signify complexity measures that exhibited a significant correlation with Stroop RT Interference ($p < 0.05$).
In light of the significant correlations between Stroop RT Interference effect and number of clauses, words, and unfilled pauses, we confirmed these findings by adding Stroop Interference as a predictor into our LMM models for these three complexity metrics. We observed a significant interaction of Condition and Stroop Interference for number of clauses ($\beta = 23.12$, $t = 2.45$, $p = 0.02$), number of words ($\beta = 156.47$, $t = 2.36$, $p = 0.02$), and number of unfilled pauses ($\beta = 28.54$, $t = 2.82$, $p < 0.01$). A subsequent simple effects analysis revealed the Stroop Interference interaction was specific to the noise condition for number of unfilled pauses ($\beta = -80.91$, $t = -2.21$, $p = 0.03$), indicating that individuals with larger interference effects (i.e., less cognitive control) produced fewer unfilled pauses when speaking in noise. Cognitive control was not predictive of number of unfilled pauses produced in silence.
Chapter 4
Discussion

Despite the prevalence of background noise in everyday language contexts, research examining the effect of noise on speech production has been dominated by studies of acoustic-related changes in production. Very little work has investigated how noise affects higher-level properties of language such as syntax. Addressing this gap in the literature, the present study examined the effect of noise on the syntactic complexity of speech production during a picture description task and asked whether differences between speaking conditions were related to individual differences in cognitive control, as measured by the AX-CPT, Flanker, and Stroop tasks. We predicted that participants speaking in noise would reduce the syntactic complexity of their speech, resulting in lower MLTU and CD and fewer T-units, clauses, and words. We predicted more disfluencies and errors due to the presumed cognitive burden of speaking in noise. We also predicted that individuals with greater cognitive control would be better able to ignore distracting auditory input and exhibit fewer differences between the noise and silence conditions.

Speakers Alter Their Speech Production in Noise

Overall, we found clear differences between speech produced in noise relative to silence. At the acoustic level, we observed that participants did speak louder in noise, which was expected given previous research examining the effect of noise on the intensity of speech (e.g., Junqua, 1993). Speakers also produced fewer clauses and words when speaking in noise. The effect of noise on number of T-units was marginal. Speaking in noisy environments is cognitively demanding, requiring focused attention on the act of speaking (Kemper et al., 2003), so it is unsurprising that participants would speak less overall.

Beyond these findings regarding the effect of noise on mean intensity and measures of
quantity, our results did not align with our predictions. Given that previous research has associated the production of mazes with utterance planning difficulty (Peach, 2013), we expected a greater number of mazes in noise. Contrary to our prediction, we found that participants produced fewer mazes in noise. One possibility is that participants are considering the needs of their listener during the picture description task, producing fewer mazes (and clauses and words) to facilitate comprehension in noisy environments. This interpretation would accord with a substantial literature on perspective-taking, the process by which individuals take into account what information their speech partner does and does not know during conversation (see Brown-Schmidt & Heller, 2018).

While this explanation is an appealing one, it is also important to note that participants produced fewer filled pauses in the noise condition as well, which seems to contradict the idea that participants were considering their listener during the task. A distinction has been drawn between filled pauses and unfilled pauses, with filled pauses argued to be listener-oriented (e.g., Clark & Fox Tree, 2002) and unfilled pauses argued to be speaker-oriented (e.g., Goldman-Eisler, 1968). Differences in the production of speaker- and listener-oriented disfluencies have been studied in individuals with Autism Spectrum Disorder (ASD), who exhibit less cognitive control (e.g., Solomon et al., 2008) and who are typically more self-centric speakers (e.g., Begeer et al., 2012). Individuals with ASD reportedly produce more unfilled pauses and fewer filled pauses than their typically-developing peers (Lake et al., 2011; Shriberg et al., 2001; but see Engelhardt et al., 2017). Taking these findings into account, it is plausible that noise would differentially affect the production of filled and unfilled pauses. If participants are considering their listener during production, we would expect them to produce more filled pauses and fewer unfilled pauses when speaking in noise, yet they produced a fewer number of both pause types.
However, the location of pauses in an utterance is an important area of further study. Previous research has proposed that pauses at clause boundaries are attributed to difficulties in utterance planning, whereas pauses that occur within clauses are attributed to lexical retrieval (de Jong, 2016). Because participants of the current study were not given time to familiarize themselves with the images before they were asked to describe them, teasing apart the effects of lexical retrieval versus utterance planning are necessary next steps towards understanding the role of pauses during production in noise.

Thus far, we have provided evidence that speakers alter their speech in the face of noise. In addition to changes in vocal intensity, participants altered other (non-acoustic) characteristics of their speech as well. Some of these findings were anticipated—participants produced fewer clauses and words in noise. Others were unexpected—participants produced fewer disfluencies (i.e., filled pauses, unfilled pauses, and mazes) in noise. In other words, speech produced in noise is more concise and less disfluent, which seems to suggest that speakers alter their speech to facilitate comprehension for their listener. However, closer inspection of listener-oriented filled pauses contradicts this interpretation; speakers produced fewer filled pauses in noise when we would expect them to produce more. To shed more light on these apparently contradictory findings, we turn to our measures of cognitive control.

Cognitive Control Modulates Speech Production in Noise

Although participants experienced inference effects for each of our cognitive control tasks (i.e., longer RTs and higher percentage of errors for AY trials compared to BX trials during the AX-CPT; longer RTs for incongruent trials compared to congruent/neutral trials during the Flanker and Stroop tasks), we only observed significant correlations for the Stroop task. We did not observe a significant relationship between differences in number of filled pauses and mazes
and Stroop interference. However, we did find significant correlations between differences in number of clauses, words, and unfilled pauses. Individuals who exhibited greater Stroop interference effects generally produced fewer clauses, words, and unfilled pauses in the noise condition, whereas those exhibiting smaller Stroop interference effects generally did not differ between the two conditions. If participants in the study were altering their speech solely for the benefit of the listener, it is counterintuitive that those with weaker cognitive control were the ones producing fewer clauses, words, and unfilled pauses in noise. Thus, it seems that participants may have been altering speech for their own benefit as well as for their listener’s benefit.

As introduced above, the challenge of speaking in noise is twofold. Not only is noise distracting for the speaker, but it also creates an adverse listening condition for the listener. When speaking in noise, speakers reduced the quantity of speech they produced as well as the number of speaker-oriented unfilled pauses, listener-oriented filled pauses, and mazes. Additionally, individuals with less cognitive control produced fewer clauses, words, and unfilled pauses in noise. Taken together, these findings suggest that there are both speaker- and listener-oriented modifications that occur when speaking in noise. Speaking in the presence of background noise entails the inhibition of distracting auditory input, so speakers may have reduced the quantity of speech and the number of unfilled pauses as they grappled with these demanding conditions. Correspondingly, we found that individuals with lower cognitive control tended to produce fewer clauses, words, and unfilled pauses in noise. Because the noise also presents a challenge for the listener, participants may have also reduced the number of mazes for their listener’s benefit as a means of facilitating comprehension. If it is the case that filled pauses assist listeners in the comprehension of complex syntactic structures (Bailey & Ferreira, 2003;
Watanabe et al., 2008), the reduction of filled pauses in noise may be reflective of reduced complexity of speech produced in noise. Alternatively, the reduction of filled pauses may be a method of streamlining speech for the listener.

**Points of Convergence and Divergence with Previous Studies**

As mentioned above, speech production in noise (beyond the acoustic level) is an understudied area, however previous work by Kemper et al. (2003) and Harmon et al. (2021) provided evidence that speakers alter certain higher-level properties of their speech due to noise. However, Kemper et al. (2003) were primarily interested in age-related differences in speech production while completing concurrent tasks, whereas Harmon et al. (2021) were interested in how different types of background noise affects speech production. In both studies, participants completed an elicitation task in which they answered prompts while speaking in noise. Most relevant to the current study, Harmon et al. (2021) similarly reported fewer unfilled pauses in noise. They proposed two potential explanations for these surprising results: speakers produced fewer unfilled pauses to maintain the listener’s focus in the face of distraction or to maintain their own focus on the task. Because speakers were exposed to noise but the listener was not, it is unclear why the reduction of unfilled pauses would be a means of maintaining the listener’s focus in the context of Harmon et al.’s (2021) study. However, both the speaker and listener were exposed to noise in our study, but given that we found that individuals with less cognitive control produced fewer unfilled pauses in noise, it is more likely that speakers reduced the number of unfilled pauses in noise to maintain their own focus on speaking without becoming distracted by the background noise.

Contrary to our findings, both Kemper et al. (2003) and Harmon et al. (2021) reported increased disfluencies in noise while we observed decreased disfluencies. However, topic of
conversation has been shown to influence the production of disfluencies (Bortfeld et al., 2001; Schachter et al., 1991), thus pronounced differences in speech elicitation techniques (open-ended questions versus a semantically-controlled picture description task) may explain these diverging results. Most crucially, unlike Kemper et al. (2003) and Harmon et al. (2021), the current study was able to examine whether speakers may alter their speech for the benefit of their listener because both the participant and the experimenter were exposed to the background noise. Based on our findings, we propose that reduced disfluencies (specifically the reduction of filled pauses and mazes), are listener-oriented speech modifications. Both Kemper et al. (2003) and Harmon et al. (2021) were not able to examine the role of the listener on production, which may also account for differences in the disfluency results. Overall, our findings strongly suggest that speakers altered their speech not only due to their own exposure to the demands of background noise, but also due to the listener’s exposure to the noise, producing both speaker- and listener-oriented speech modifications. We consider the reduction of clauses, words, and unfilled pauses to be speaker-oriented because they are associated with cognitive control.

**Qualitative Assessments of Syntactic Complexity**

Although speakers exhibit clear differences between noise and silence, several open questions remain. One of these questions surrounds what strategies speakers used to reduce their quantity of speech in noise. Although they were speaking less, were these descriptions qualitatively worse in noise or were they better and more efficient? Did participants alter the syntactic structures they produced? Our two most widely utilized metrics of syntactic complexity, MLTU and CD, did not yield significant differences between the noise and silence conditions. These results were surprising given that we expected participants to produce simpler structures in noise (i.e., shorter MLTU and lower CD). However, we note that MLTU and CD
are coarse measures that are unable to inform us about the types of structures participants produced. Thus, a more qualitative assessment of the production data may reveal differences in syntactic complexity. Future analyses of the data will examine whether participants differ in their production of structures that are considered to be more complex, such as those involving non-canonical word orders (e.g., passive sentences) and long-distance dependencies (e.g., object relative clauses). Another interesting structure that may warrant further attention is subject doubling. Subject doubling occurs when the subject of a clause appears twice, often in the form of a full noun phrase followed by a pronoun (e.g., *the girl walking towards the sandcastle* [*she* has her hair up in a half ponytail]). While subject doubling is generally uncommon in English, it is a feature of a variety of Canadian English spoken in Kapuskasing, Ontario (Tagliamonte & Jankowski, 2019) and has been reported in Southern United States English as well (Southard & Muller, 1993). Within the context of the current study, a question of interest is why speakers produced double subjects. Is it simply a dialectal feature or might it be a means of offloading potential cognitive burden associated with introducing a new subject and producing a full noun phrase? If it latter is true, we would expect to see more instances of subject doubling in the noise condition.

Additionally, the process of segmenting the transcriptions into T-units proved to be more challenging than anticipated. For example, there were several instances where participants embedded a full T-unit into another T-unit. Consider the following example: *and laying on the camel she has her back on the camel and it looks like she’s asleep* is a woman with long hair. The inserted T-unit (in this case, two T-units) within the initial T-unit is bolded. In these cases, we did not exclude the embedded T-units from analyses because the embedded T-units themselves were grammatical and the removal of the embedded T-units does not affect the
grammaticality of the T-unit in which they were embedded. However, these embedded T-units could also be considered an interruption that occurred within the utterance planning process, so it remains an open question how to appropriately analyze these constructions.

Another point of contention surrounded how to treat expressions of uncertainty (e.g., *it seems/looks/appears; I think/guess/suppose/assume*). The use of words or phrases to convey uncertainty, possibility, and probability is referred to as hedging, and this behavior has been argued to be characteristic of women’s speech (e.g., Holmes, 1990; Mondorf, 2002). Considering that ninety percent of our participants were female, this may explain the prevalence of hedges in the production data. Hedges that include verbs such as *seem, look, appear,* and *guess* can take sentential (sometimes nominal) complements, so when these hedges were followed by sentential complements (e.g., *it seems that the castle in the top corner is destroyed*), these expressions are more syntactically integrated. However, participants also produced hedges in positions where the verb was not followed by sentential complements (e.g., *she’s petting a seal looking thing I’m assuming; there’s I guess a wannabe prince*). In these cases, the hedges could be construed as fillers and analyzed as disfluencies. We did not code hedges as disfluencies, and hedges were included in our analyses of quantity and MLTU because they still conveyed relevant information to the listener (i.e., the speaker’s uncertainty regarding their descriptions).

**Conclusion**

The present study examined whether (1) background noise affects the syntactic complexity of speech production; and (2) cognitive control predicted noise-induced differences in complexity. Despite some limitations, we found compelling evidence that speakers alter the way they speak in noise versus silence. Due to the nature of online data collection, the level of background noise during the picture description task likely varied across participants. Moreover,
we acknowledge the value in conducting this study under more naturalistic, in-person conditions. Nonetheless, the results of the present study suggest that speakers take into account the needs of their listener when speaking in noisy environments while also altering speech for their own benefit. We found evidence that speakers produced fewer mazes and filled pauses in noise, which could be indicative of speakers streamlining their speech for the sake of their listener. Speakers also produced fewer clauses, words, and unfilled pauses, which may be indicative of speakers simplifying speech to reduce their own cognitive burden. Most notably, participants with lower cognitive control, as indexed by larger Stroop RT interference effects, produced fewer clauses, words, and unfilled pauses in noise, providing further support that these reductions were for the speaker’s benefit. To further examine the distinction between speaker- and listener-oriented speech modifications, future analyses will consider more qualitative assessments of the production data, examining the overall quality of the descriptions produced in noise versus silence and potential differences in the types of syntactic structures produced in each condition.
REFERENCES


https://doi.org/10.2307/1165818


https://doi.org/10.1121/1.405631


https://doi.org/10.1016/S0167-6393(96)00041-6


https://doi.org/10.1080/13825580500501496

https://doi.org/10.18637/jss.v082.i13


https://doi.org/10.1109/ICASSP.1985.1168217


Tagliamonte, S. A., & Jankowski, B. L. (2019). Grammatical convergence or microvariation?


APPENDIX A

Full Summary of Statistical Analyses

Table A1. Coefficients and corresponding $t$-values and $p$-values for each predictor in a linear mixed effects model examining all main effect and interactions of Condition and Run on eleven measures of complexity. Significant values are bolded.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coefficient</th>
<th>$T$-value</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MODEL 1: Mean Length of T-unit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td>0.10</td>
<td>1.02</td>
<td>0.31</td>
</tr>
<tr>
<td>Run</td>
<td>0.09</td>
<td>1.03</td>
<td>0.30</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>−0.05</td>
<td>−0.49</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>MODEL 2: Clausal Density</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td>0.02</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td>Run</td>
<td>0.01</td>
<td>0.92</td>
<td>0.36</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>−0.003</td>
<td>−0.21</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>MODEL 3: Number of T-units</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td>0.59</td>
<td>1.76</td>
<td>0.08</td>
</tr>
<tr>
<td>Run</td>
<td>0.47</td>
<td>1.55</td>
<td>0.12</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>0.34</td>
<td>1.04</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>MODEL 4: Number of Clauses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td><strong>1.47</strong></td>
<td><strong>3.25</strong></td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>Run</td>
<td><strong>1.41</strong></td>
<td><strong>2.37</strong></td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>0.31</td>
<td>0.73</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>MODEL 5: Number of Words</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td><strong>10.94</strong></td>
<td><strong>3.43</strong></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Run</td>
<td><strong>9.49</strong></td>
<td><strong>2.30</strong></td>
<td><strong>0.03</strong></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>2.26</td>
<td>0.76</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>MODEL 6: Number of Unfilled Pauses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)</td>
<td><strong>2.59</strong></td>
<td><strong>4.30</strong></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Run</td>
<td><strong>1.34</strong></td>
<td><strong>2.51</strong></td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>Model</td>
<td>Condition (Noise vs. Silence)</td>
<td>Run</td>
<td>Condition (Noise vs. Silence)*Run</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------</td>
<td>-----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><strong>Total Pause Duration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.80</td>
<td>1.84</td>
<td>0.07</td>
</tr>
<tr>
<td>Run</td>
<td>0.48</td>
<td>1.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>–0.02</td>
<td>–0.04</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Number of Filled Pauses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.79</td>
<td>3.38</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Run</td>
<td>0.258</td>
<td>1.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>0.62</td>
<td>2.68</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td><strong>Number of Mazes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.63</td>
<td>2.54</td>
<td>0.01</td>
</tr>
<tr>
<td>Run</td>
<td>0.24</td>
<td>1.09</td>
<td>0.28</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>0.12</td>
<td>0.50</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Number of Errors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.02</td>
<td>0.17</td>
<td>0.87</td>
</tr>
<tr>
<td>Run</td>
<td>0.13</td>
<td>1.12</td>
<td>0.27</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>0.05</td>
<td>0.37</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>Number of Omitted Words</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.03</td>
<td>0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>Run</td>
<td>0.20</td>
<td>1.91</td>
<td>0.06</td>
</tr>
<tr>
<td>Condition (Noise vs. Silence)*Run</td>
<td>&lt; –0.0001</td>
<td>–0.001</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table A2. Correlation coefficients and p-values for the non-significant correlation analyses between Silence–Noise differences for each complexity measure and cognitive control measure.

<table>
<thead>
<tr>
<th>Cognitive Control Measure</th>
<th>Complexity Measure</th>
<th>Correlation Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AX-CPT Behavioral Shift Index</td>
<td>Number of Clauses</td>
<td>−0.12</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Number of Words</td>
<td>−0.18</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Number of Unfilled Pauses</td>
<td>−0.19</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Number of Filled Pauses</td>
<td>−0.003</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Number of Mazes</td>
<td>0.15</td>
<td>0.27</td>
</tr>
<tr>
<td>Flanker RT Interference</td>
<td>Number of Clauses</td>
<td>0.06</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Number of Words</td>
<td>0.11</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Number of Unfilled Pauses</td>
<td>0.13</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Number of Filled Pauses</td>
<td>0.09</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Number of Mazes</td>
<td>−0.25</td>
<td>0.10</td>
</tr>
</tbody>
</table>
APPENDIX B

Transcription Sample

The transcription below is taken from Participant 8 describing the beach scene image in silence. Each T-unit is numbered. The code designating each clause type has been inserted at the end of each T-unit in brackets. All words enclosed in parentheses were removed prior to calculating MLTU, CD, and number of T-Units, clauses, and words.

(1) it looks like they're (pause: 1200 ms) sort of on a beach [MC] [NOM]

(2) but (there’s gra#) it looks like there’s grass on the left side with (a house#) (pause: 929 ms) almost like a little beach house with a garage and like two windows on the side and then one window and then one door (pause: 892 ms) [MC] [NOM]

(3) it’s like a one story house [MC]

(4) there’s a mountain in the background (pause: 408 ms) or like a (pause: 920 ms) higher landmark (on the#) in the background [MC]

(5) there’s a guy (pause: 669 ms) flying a kite (pause: 1812 ms) [MC] [PRT]

(6) and (i#) $there’s clouds in the sky [MC]

(7) ($/it) doesn’t look like there’s any sun (pause: 854 ms) [MC] [NOM]

(8) (there#) (pause: 585 ms) in the bottom left_hand corner (there lo#) there looks like a soccer ball [MC]

(9) but I don’t know (pause: 1069 ms) why it’s just by itself just rolling away it looks like it (pause: 1277 ms) [MC] [NOM] [PRT]

(10) and then above it (pause: 576 ms) there’s a (pause: 320 ms) girl petting a sea_lion (pause: 1031 ms) [MC] [PRT]

(11) and she’s underneath an umbrella (pause: 558 ms) on a towel [MC]
she’s in a dress and looks like (pause: 567 ms) just slides for shoes [MC]

she has short hair (pause: 1031 ms) [MC]

she’s (pause: 1421 ms) getting a drink from (pause: 622 ms) a younger (pause: 1050 ms) it looks like (pause: 389 ms) (&like) a boy (pause: 508 ms) [MC]

and he has like a (beach_ball#) (pause: 313 ms) beach volleyball I think in his hand or a beach_ball (pause: 799 ms) [MC]

he has short hair as well [MC]

and he looks like he has like tennis_shoes on (pause: 863 ms) [MC] [NOM]

and (pause: 1523 ms) (UH) in about the middle (pause: 520 ms) bottom portion of the picture (there’s another#) there’s two people that look like they're (pause: 567 ms) building sandcastles [MC] [REL] [NOM]

but they look like they’re arguing (&arguing) on (pause: 864 ms) where everything goes (pause: 863 ms) [MC] [NOM] [NOM]

the one on the right’s a little bit younger [MC]

and (pause: 1180 ms) they’re wearing a hat with a (pause: 455 ms) dress and it looks like leggings (pause: 399 ms) [MC]

and the other one has dreadlocks with a tank_top and shorts (pause:754 ms) [MC]

and there are seashells around them (pause: 2331 ms) [MC]

(UH) (pause: 687 ms) and then to the right (pause: 437 ms) there’s another person walking with (a#) (pause: 360 ms) it looks like a pail or a bucket (pause: 1515 ms) [MC] [PRT]

and their hair is (pause: 647 ms) down but tied up a little bit (pause: 719 ms) in the back [MC]
and they have (pause: 913 ms) (UH short sl#) (pause: 483 ms) a t-shirt and shorts (pause: 879 ms) [MC]

and it looks like when you get closer to the water there might be more rocks (pause: 399 ms) cause there might be like a pathway before it reaches the water (pause: 949 ms) [MC] [ADV] [NOM] [ADV] [ADV]

and (pause: 727 ms) there’s a mermaid (pause: 550 ms) on a rock [MC]

and she looks like she’s either drinking something or she’s gonna play something do something with her mouth (pause: 727 ms) [MC] [NOM] [NOM] [INF] [INF]

and she has it looks like a t-shirt on (pause: 647 ms) as well [MC]

and then there’s a (little#) (pause: 951 ms) like little sailboat in the water [MC]

but it looks like (pause: 879 ms) someone like made it out of like paper (pause: 442 ms) [MC] [NOM]

it’s not a life_size (pause: 951 ms) (UM) (pause: 790 ms) boat (pause: 514 ms) [MC]

and then to the right of that (there’s a#) (pause: 381 ms) it looks like a (&like a) sea_monster (pause: 835 ms) [MC]

it looks like a serpent in a way (pause: 645 ms) with weird eyes and pointy teeth (pause: 642 ms) [MC]

and it’s very long [MC]

and (pause: 1178 ms) besides that and the little sailboat there’s nothing really else in the water [MC]

there’s the (pause: 374 ms) mermaid [MC]

($/it) looks like she’s about to go into the water (pause: 796 ms) [MC] [NOM] [INF]

but she’s just sitting there (pause: 265 ms) on the rock [MC]