ASSESSING THE COGNITIVE LOAD OF TALKING IN WHISPERS AND CLEAR SPEECH WITH A MODIFIED SIMON TASK

Keiko Ishikawa[1*] **Christy Kirkham**[1] **Julia Samenuk**[1] **Rachel Weaver**[1] **Diana Orbelo**[2]
[1] Department of Communication Sciences and Disorders, University of Kentucky, Lexington, KY, USA
[2] Mayo Clinic, Department of Otolaryngology-Head and Neck Surgery, Rochester, MN, USA

ABSTRACT

This study aimed to examine the cognitive demand associated with different speech production modes. Fifteen vocally healthy college students performed a Simon task, a method widely used to evaluate cognitive load via response times to visual stimuli. Participants performed the task both in silence and while counting numbers in three speech modes: habitual speech, whisper, and clear speech. The reaction time was measured for 'congruent' and 'incongruent' trials, which referred to whether the locations of the square and the key assigned to the square's color matched or mismatched, respectively. Results revealed that all modes of speech production significantly increased cognitive load compared to silence. Among the speech production modes, clear speech elicited longer reaction times than whispering or habitual speech. However, this effect was only observed in congruent trials, suggesting that the cognitive load, associated with managing spatial incongruity, might mask subtle differences in cognitive load between speech production modes. Collectively, our findings suggest that clear speech is more cognitively demanding than habitual speech or whispering, and that reaction times in the Simon task may serve as an effective proxy for cognitive load during speech production.

Keywords: *cognitive load, speech modification, dual task*

1. INTRODUCTION

The success of skill transfer in voice and speech therapy hinges on the development of automaticity, or the ability to execute tasks with minimal cognitive resources [1], which is crucial for the effective application of a learned technique in real-world settings. Patients often struggle to apply learned techniques in real-life communication, potentially due to insufficient development of automaticity. Clinicians typically assess automaticity through patient performance. Such performance-based assessments do not capture the patient's mental effort for the performance. As such, this measure could potentially fail to identify cases in which a successful performance is achieved at the cost of an unsustainable cognitive load. Factoring cognitive load into the evaluation of automaticity may provide a more comprehensive understanding of a patient's capability to effectively apply learned techniques.

Clear speech and whispering represent two contrasting speech modification techniques [2] that require intentional alterations to speech motor programs. These modifications were executed for opposing reasons: clear speech aims to enhance intelligibility, while whispering aims to reduce it. Clear speech, in particular, has been incorporated into speech therapy due to its proven benefits in intelligibility. While both techniques potentially increase cognitive demand compared to habitual speech, the extent of this cognitive load remains largely unexplored. Quantifying this cognitive load has direct clinical implications, as it could indicate the degree of difficulty involved in developing automaticity. For instance, if a specific speech modification technique requires less cognitive load than others, it suggests that patients may find it easier to achieve automaticity with that technique.

The dual-task paradigm is an experimental approach designed to assess cognitive load by evaluating the interference between primary and secondary tasks [3, 4]. This paradigm has been used to study the effects of concurrent tasks on speech motor performance [5-10]. Generally, these studies assume that secondary tasks increase

[*] ***Corresponding author***: *Ishikawa@uky.edu*

the cognitive load, thus impacting speech performance. However, these paradigms inherently do not measure the specific cognitive load associated with the speech modification [11, 12]. Recognizing this limitation underscores the need for alternative strategies that can accurately quantify the cognitive load specific to speech modification.

The Simon task is a widely used tool for assessing cognitive load by measuring response time to specific stimuli [13-14]. This task typically presents a stimulus, such as a colored square, in a specific location on a screen. It requires participants to respond based on the feature of the stimulus (e.g., color), while ignoring the other (e.g., location). Response time and accuracy are often influenced by the congruency between the location of a stimulus and the response button, an effect known as the Simon effect. Reaction times are usually faster and responses more accurate in congruent trials compared to incongruent ones, indicating the additional cognitive load imposed by managing spatial incongruity. Given these characteristics, the Simon task could potentially serve as a valuable tool to measure the cognitive load associated with speech modification. The aspect of congruency in the Simon task provides an opportunity to evaluate the appropriate level of difficulty in a secondary task that is necessary to accurately detect the cognitive load associated with speech modification.

In response to the gap in our understanding of cognitive load in speech modification, this study utilized the Simon task to estimate the cognitive load associated with specific speech modification techniques, namely clear speech and whispering. Two hypotheses were tested: 1) Speech modification techniques would impose a greater cognitive load than habitual speech, and 2) The cognitive load required to manage spatial incongruity would influence the Simon task's sensitivity in detecting differences in cognitive load associated with speech modifications.

2. METHOD

The experimental protocols for this study were approved by the Institutional Review Board of the University of Kentucky. All participants provided written informed consent to participate in the experiment.

Participants: The participants in the study were ten female and one male native speakers of American English, between 19 and 23 years of age. None of the participants had a history of speech-language or hearing disorders. None of them had vision or physical conditions that interfered with the task.

Task and Procedure: Participants were seated in a quiet room facing a computer monitor. Stimuli were presented on the monitor using Inquisit 6 software (Millisecond Software). Each trial started with an 800-ms fixation cross in the center of the screen, followed by a 250-ms blank interval. A blue or red square then appeared on the left or right side of the screen for 1,000 ms, and participants were instructed to respond as quickly and accurately as possible by pressing the left 'A' key for a blue square and the right 'L' key for a red square. When the color of the square and the location of the response key matched, the trial was deemed as 'congruent'. Conversely, when there was a mismatch between the color of the square and the location of the key, the trial was deemed as 'incongruent'. To prevent previous trials from influencing participants, a 500-ms blank interval separated each trial: see Fig. 1.

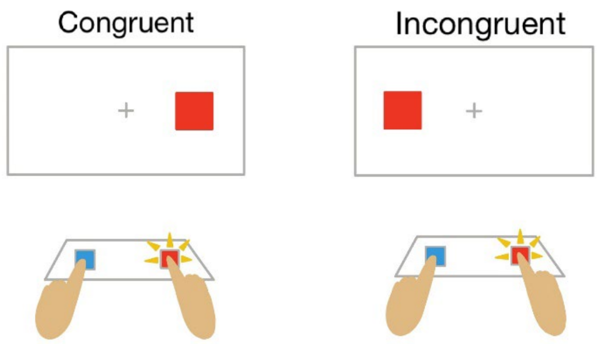

Figure 1. Description of the Simon Task

Participants were given a practice trial before the experimental trial. During the practice trial, participants counted numbers aloud in habitual speech while performing the Simon task. Participants proceeded to the experimental trial only after achieving 100% accuracy in the practice. Participants completed four separate experimental blocks of the Simon task in silence or counting numbers in habitual speech, whisper, and clear speech. The order of the speech conditions was counterbalanced across participants to control for order effects. In each block, participants completed 28 trials, 14 congruent and 14 incongruent trials, presented in random order. The experimenter monitored participants' speech throughout the task to ensure they maintained the instructed speech mode.

Statistical Analyses: A two-way repeated measures ANOVA with the within-subject factors of speech style (silent, habitual, whisper, clear) and congruence (congruent, incongruent) was conducted. The dependent variable was reaction time.

3. RESULTS

The average reaction time and the standard errors (SE) for each speech style and congruence condition are presented in Table 1. The results of the two-way repeated measures ANOVA indicated a significant main effect of speech style, $F(3, 14) = 45.87$, $p < .001$, $\eta^2 = .076$, as well as a significant main effect of congruence, $F(1, 14) = 6.60$, $p = .010$, $\eta^2 = .004$: see Fig 2.

Table 1. Descriptive Statistics of Average Reaction Time and Standard Error.

Task	Congruence	N	Reaction Time (ms)	SE
Silent	Congruent	210	423.16	8.21
Silent	Incongruent	210	444.88	6.83
Habitual	Congruent	210	500.38	12.18
Habitual	Incongruent	210	519.49	11.39
Whisper	Congruent	210	475.20	9.61
Whisper	Incongruent	210	529.36	13.56
Clear	Congruent	210	572.48	17.67
Clear	Incongruent	210	559.59	12.94

A series of pairwise t-tests with Bonferroni correction were conducted to compare the reaction time between different task conditions within congruent and incongruent trials. In the congruent trials, there were significant differences between Silent and Habitual speech, $t(209) = -6.08$, $p_{adj} < .001$, Silent and Whisper, $t(209) = -4.56$, $p_{adj} < .001$, and Silent and Clear speech, $t(209) = -8.24$, $p_{adj} < .001$. Additionally, there were significant differences between Habitual and Clear speech, $t(209) = -3.50$, $p_{adj} = .003$, and Whisper and Clear speech tasks, $t(209) = -5.17$, $p_{adj} < .001$. No significant differences were found between Habitual and Whisper, $t(209) = 1.90$, $p_{adj} = .355$.

In the incongruent trials, significant differences were observed between Silent and Habitual speech, $t(209) = -6.13$, $p_{adj} < .001$, Silent and Whisper, $t(209) = -5.87$, $p_{adj} < .001$, and Silent and Clear speech, $t(209) = -8.36$, $p_{adj} < .001$. No significant differences were found between Habitual and Whisper, $t(209) = -0.569$, $p_{adj} = 1.00$, Habitual and Clear speech, $t(209) = -2.55$, $p_{adj} = .069$, and Whisper and Clear speech, $t(209) = -1.68$, $p_{adj} = .571$.

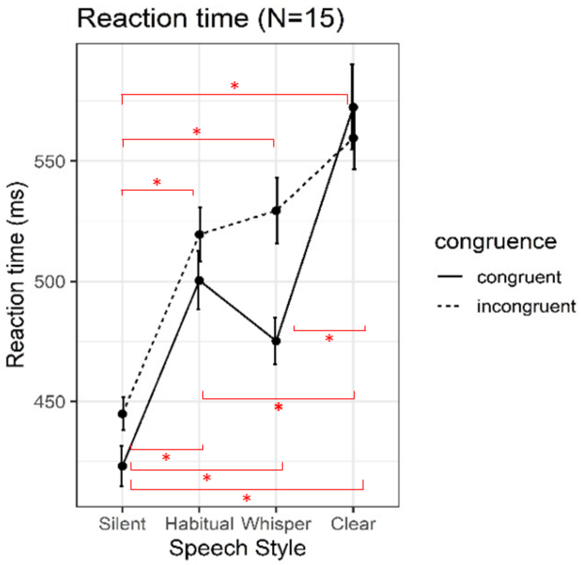

Figure 2. Line plot showing the average reaction time for four speech production styles in congruent and incongruent trials. The error bars indicate standard error. The asterisks indicate statistically significant differences between the conditions.

4. DISCUSSION

4.1 General discussion

The present study aimed to investigate the cognitive load associated with two speech modification techniques, whispering and clear speech, using the Simon task. Two hypotheses were tested: 1) Speech modification techniques would impose a greater cognitive load than habitual speech, and 2) The cognitive load required to manage spatial incongruity would influence the Simon task's sensitivity in detecting differences in the cognitive load associated with speech modifications.

The results partially supported the hypotheses. In congruent trials, whispering did not significantly increase the reaction time compared to habitual speech, suggesting that it did not impose additional cognitive load. Conversely, clear speech significantly increased the reaction time in comparison to both habitual speech and whispering, implying it imposed additional cognitive load. These results may suggest that whispering is a relatively automatic process that does not necessitate substantial changes to the speech motor program. In contrast, clear speech may not be as automatic and may require more considerable adjustments to the speech motor program, potentially making it more challenging for individuals to apply consistently and effectively. However, the effects of the speech production modes were not observed in the incongruent trials.

Findings in the congruent trials align with those reported by Whitfield et al., who demonstrated that higher effort speaking styles, such as clear speech, require greater attentional resources than habitual speech [12]. It's worth noting that a longer reaction time is typically interpreted as the task requiring more intensive processing resources or the engagement of more extensive or complex neural circuits [15]. Accordingly, the observed increase in response time for clear speech in our study could suggest that it engaged more extensive areas of the motor cortex, which might have diverted resources from performing the Simon task. Moreover, maintaining clear speech likely necessitates ongoing auditory monitoring to ensure the quality of speech sounds, thereby potentially requiring greater activation of the auditory cortex compared to other speech styles. It's also plausible that the subsequent adjustments to the speech motor program based on the auditory input could necessitate more extensive activation of the prefrontal cortex, given its role in executive functioning. While the significant difference in reaction times between whispering and clear speech modes was observed only in congruent trials, it is possible that the shorter reaction time associated with whispering reflects reduced activity in these neural regions. However, the underlying mechanisms behind these differences remain unclear and merit further investigation.

The reaction time differences observed in this study align with those reported in prior research involving speech tasks. For instance, Hsieh et al. examined reaction times to visual events during no conversation, covert (i.e., thinking), and overt (i.e., verbalizing) conversation tasks [16]. Their results showed that compared to no conversation, covert conversation increased the mean visual event reaction time by 41 ms, and overt conversation increased it by 73 ms. Similarly, in this study, counting aloud in habitual speech increased the mean reaction time by 77 ms, compared to no counting. The reaction time further increased by 72 ms when participants were engaged in clear speech compared to habitual speech, and by 97 ms from whispering to clear speech. While both studies highlight the cognitive load associated with speech production, it is essential to acknowledge the differences in the nature of the speech tasks. The speech task from Hsieh et al. required the generation of ideas and language, whereas the speech task of our study did not. Given the limited cognitive demand for content and language generation in our study, the observed increase in reaction times likely reflects the recruitment of greater neural networks required for speech modifications.

The longer reaction time for the incongruent trials corroborates findings of previous studies that utilized the Simon task [14]. Incongruent trials showed significant differences in the reaction time between silence and all speech production modes but did not reveal significant differences between the speech production modes. The lack of differences may suggest that the cognitive resources allocated to manage the spatial incongruency was greater than the resource allocated to speech modification. The results also suggest that a simpler secondary task, which does not consider the effect of congruency on cognitive load, would be more suitable for detecting the cognitive load associated with speech modification.

4.2 Limitations and future directions

Several limitations should be considered when interpreting the results. First, the sample size is relatively small, with only ten female and one male participant, which may limit the generalizability of the findings. Moreover, the participants' age range was narrow, between 19 and 23 years old, and all were native speakers of American English. Another limitation is the use of a single task to estimate cognitive load. While the Simon task is a well-established experimental method, incorporating psychophysiological measures of cognitive load, such as pupillometry and skin conductance, could provide more robust and convergent evidence. Furthermore, the study only investigated two speech modification techniques—clear speech and whispering. Including other speech modification techniques or strategies in future studies would allow for a more comprehensive understanding of the cognitive demands associated with real-world communication. Lastly, the study relied on the participants' abilities to maintain the instructed speech mode, a factor subject to individual variability and compliance. The incorporation of objective measures to confirm and quantify the speech mode, such as acoustic analysis of speech, would aid in evaluating the speaker's adherence to the instructed speech techniques.

While counting aloud was chosen to minimize the potential interference from language generation, its ecological validity is limited for real-world communication contexts. Future studies could consider using more ecologically valid speech tasks, such as reading a passage or engaging in conversation, to better understand the cognitive demands of speech modification techniques in real-world situations. These tasks could involve more complex language processing and social interaction, which may better reflect the challenges faced by individuals with speech and communication disorders in their daily lives. Furthermore, using a functional neuroimaging method could help to identify the specific cognitive processes and neural mechanisms involved in different speech modification techniques. Such information would provide a more comprehensive understanding of their role in hindering or facilitating skill transfer in voice and speech rehabilitation.

5. CONCLUSION

This study utilized a dual-task paradigm to enhance our understanding of the cognitive load associated with speech modification techniques. Our findings indicate that the cognitive load depends on the technique: clear speech requires more cognitive resources compared to both habitual speech and whispering. Furthermore, the cognitive load incurred by the secondary task (i.e., the Simon task) should be considered when investigating the cognitive aspects of speech modification. These findings suggest that continued research in this area holds the potential to optimize individual strategies for speech modification and ultimately improve communication outcomes for individuals with voice and speech disorders.

6. ACKNOWLEDGMENTS

The authors would like to express their appreciation to Olivia Stevens and Corinne Wing for their assistance with data collection. This study was funded by start-up funds awarded to the principal investigator (Ishikawa) by the University of Kentucky.

7. REFERENCES

[1] A. M. Haith and J. W. Krakauer, "The multiple effects of practice: skill, habit and reduced cognitive load," *Curr. Opin. Behav. Sci.*, vol. 20, pp. 196-201, 2018.

[2] R. M. Uchanski, "Clear speech," *in The Handbook of Speech Perception*, pp. 207-235, 2005.

[3] K. E. DeLeeuw and R. E. Mayer, "A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load," in *Journal of Educational Psychology*, vol. 100, no. 1, pp. 223, 2008.

[4] J. Sweller, P. Ayres and S. Kalyuga, "Measuring cognitive load," in *Cognitive Load Theory*, pp. 71-85, 2011.

[5] C. Dromey and A. Benson, "Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance," in *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 5, pp. 1234-1246, 2003.

[6] C. Dromey and E. Shim, "The effects of divided attention on speech motor, verbal fluency, and manual task performance," in *Journal of Speech, Language, and Hearing Research,* vol. 51, no. 5, pp. 1171-1182. 2008.

[7] C. Dromey and K. Simmons, "Bidirectional interference between simulated driving and speaking," in *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 7, pp. 2053-2064, 2019.

[8] K. Bunton and C. K. Keintz, "The use of a dual-task paradigm for assessing speech intelligibility in clients with Parkinson disease," in *Journal of Medical Speech-Language Pathology*, vol. 16, no. 3, pp. 141, 2008.

[9] M. K. MacPherson, "Cognitive load affects speech motor performance differently in older and younger adults," in *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 5, pp. 1258-1277, 2019.

[10] D. J. Bailey and C. Dromey, "Bidirectional interference between speech and nonspeech tasks in younger, middle-aged, and older adults," in *Journal of Speech, Language, and Hearing Research*, vol. 58, no. 6, pp. 1637-1653, 2015.

[11] J. A. Whitfield, Z. Kriegel, A. M. Fullenkamp and D. D. Mehta, "Effects of concurrent manual task performance on connected speech acoustics in individuals with Parkinson disease," in *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 7, pp. 2099-2117, 2019.

[12] J. A. Whitfield, S. R. Holdosh, Z. Kriegel, L. E. Sullivan and A. M. Fullenkamp, "Tracking the Costs of Clear and Loud Speech: Interactions Between Speech Motor Control and Concurrent Visuomotor Tracking," in *J Speech Lang Hear Res*, vol. 64, no. 6s, pp. 2182-2195, 2021.

[13] R. W. Proctor, "Playing the Simon game: Use of the Simon task for investigating human information processing," in *Acta Psychologica*, vol. 136, no. 2, pp. 182-188, 2011.

[14] J. R. Simon, "The effects of an irrelevant directional cue on human information processing," in *Advances in Psychology* (Elsevier), pp. 31-86, 1990.

[15] X. Liu, M. T. Banich, B. L. Jacobson, & J. L. Tanabe, “Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI.” *Neuroimage*, vol. 22, no. 3, pp. 1097-1106, 2004.

[16] L. Hsieh, R. A. Young, S. M. Bowyer, J. E. Moran, R. J. Genik II, C. C. Green, ... & S. Seaman. “Conversation effects on neural mechanisms underlying reaction time to visual events while viewing a driving scene: fMRI analysis and asynchrony model.” in *Brain Research*, 1251, pp. 162-175, 2009.