DOI: 10.61782/fa.2023.0929

THE USE OF PUPILLOMETRY TO STUDY MENTAL DEMAND DURING SPEECH PRODUCTION IN NOISE

Keiko Ishikawa[1*] **Elisabeth Coster**[2] **Hannah Li**[2] **Diana Orbelo**[3]
[1] Department of Communication Sciences and Disorders, University of Kentucky, USA
[2] Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, USA
[3] Mayo Clinic, Department of Otolaryngology-Head and Neck Surgery, USA

ABSTRACT

Understanding the cognitive demand required for implementing speech therapy techniques in real-world environments has significant clinical implications. This study used self-rating and pupillometry to investigate the mental demand associated with speaking differently in various noise conditions. Five speakers with healthy voices read aloud in casual and clear speech in quiet conditions and while listening to multi-talker (MT) noise, reversed two-talker (RevMT) noise, and speech-shaped (SS) noise. The speakers rated their perceived mental demand after each task using a modified NASA-TLX scale, while the change in their pupil size was monitored using an eye tracker. The results of this self-rating revealed that mental demand was highest in the MT noise condition, followed by RevMT noise. In contrast, SS noise did not significantly increase mental demand. Pupillometry data showed that the pupil dilation was greatest for the MT noise, for both casual and clear speech, corroborating the self-rating data. Interestingly, the dilation for clear speech was greater than for casual speech in quiet, but less in RevMT and SS noises. These findings illustrate a complex interaction between speech modification and acoustic environments and suggest the potential of pupillometry as an objective measure for quantifying mental demand in speech production.

Keywords: *cognitive load, speech modification, speech production, background noise, pupillometry*

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

1. INTRODUCTION

Speech training and therapy often require individuals to modify their speaking patterns to enhance intelligibility, particularly for those with voice and speech production disorders. During therapy, patients learn new techniques to improve their speech. The ultimate goal is skill transfer, enabling them to develop automaticity with the technique and apply it in various settings outside the controlled training environment [1]. As individuals achieve automaticity, they require fewer cognitive resources to perform tasks, specifically, utilizing learned speech techniques [2]. Real-world communication environments are often filled with distractions, including background noise, which can affect speech production [3] and cognitive load [4-6]. Developing automaticity in these environments, where distractions are abundant, is crucial for ensuring that individuals can effectively use their learned speech techniques in everyday life.

Because background noise affects speech production and cognitive function, it potentially impacts the transfer process of voice therapy. Speech perception research has traditionally classified noises into energetic masking noise [7], which physically masks speech signals, and informational masking noise, which contains linguistic information that competes with target speech. It has been shown that informational masking noise requires greater cognitive load for speech perception than energetic masking noise [8]. How these different types of noise affect cognitive load during speech production is unclear. If noise distracts an individual from a speech production technique, they must pay more attention to the technique, potentially increasing the cognitive load.

Clear speech is a speech modification technique aimed at enhancing intelligibility by deliberately overemphasizing speech sounds [9]. It involves a slower speaking rate, increased pauses, greater articulatory precision, and expanded vowel space [10-12]. These characteristics help listeners better distinguish speech sounds, making it easier to understand the speaker. Clear speech has been shown to improve intelligibility as much as 30% [13]. It has been assumed that clear speech is relatively easy to produce thus it requires low cognitive load [9]. This assumption has made clear speech an attractive choice for incorporation into a voice therapy program [14].

Estimating the cognitive load associated with performing a task, such as using a learned speech technique, can provide insight into how well the automaticity has been developed. However, this estimation is challenging for several reasons. Firstly, speech performance alone does not indicate the level of automaticity development, as individuals can improve performance by concentrating more on the task. Secondly, although rating scales could be used to estimate the load, subjective judgments are often unreliable [15]. Consequently, there is a need for an objective, physiological measure of cognitive load associated with speech production to overcome these limitations.

Pupillometry is a widely used technique in cognitive psychology for estimating cognitive load, attention, and emotional arousal [16-17]. It has also been extensively employed in speech perception studies to estimate listening effort [18]. It involves measuring the changes in pupil size, which can be influenced by various cognitive processes. The physiological underpinnings of pupillary dilation are related to the autonomic nervous system, particularly the balance between the sympathetic and parasympathetic nervous systems. As cognitive load, attention, or emotional arousal increases, the sympathetic nervous system becomes more active, leading to an increase in pupil size. Conversely, a decrease in these psychological constructs results in the activation of the parasympathetic nervous system, causing the pupil to constrict [17]. By examining the fluctuations in pupil diameter, researchers can gain insights into the cognitive and emotional states of an individual, making pupillometry a valuable tool for studying a variety of psychological phenomena, including speech perception and listening effort.

Despite its potential, pupillometry has not been widely used to measure cognitive load during speech modification in noisy environments. Therefore, this study aimed to investigate the feasibility of using pupillometry to measure the cognitive effort associated with speech modification in the presence of noise. Specifically, we examined the effects of different types of masking noise (informational versus energetic) and speech production techniques (casual versus clear speech) on cognitive load.

2. METHODS

The experimental protocols for this study were approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign. All participants provided informed consent.

2.1 Preparation of Noise Stimuli

Three types of noise stimuli were prepared: multi-talker noise (MT noise), reversed multi-talker noise (RevMT noise), and speech-shaped noise (SS noise). Two native speakers of American English recorded sentences from the BKB sentence lists. The intensity of the recordings was standardized to 60 dB SPL using PRAAT.

2.2 Speech Production Experiment

Speaker Participants: Five native speakers of American English aged 19-25 participated in the study. None had a history of voice, speech, or hearing disorder. Their vocal status was screened by a licensed speech-language pathologist.

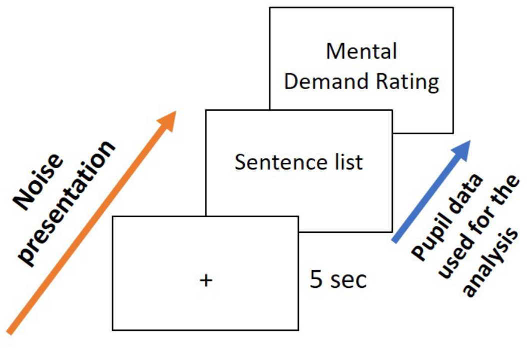

Figure 1. Schematic description of the trial

Study Materials: The stimuli consisted of a set of three slides for each noise condition. The first slide displayed a baseline cross for five seconds. The second slide displayed instructions for the speaking style and reading materials, which included two lists of ten sentences from the Hearing in Noise test [19]. The last slide was a self-report survey asking participants to rate the degree of mental effort on a 20-point scale, modeled after the NASA-TLX scale [20], which assesses perceived workload across various tasks, including mental demand: see Fig. 1.

Procedures: Participants underwent training to ensure they could produce casual and clear speech. The experiments were conducted in a soundproof booth. Participants sat in front of a computer monitor displaying the stimuli. An eye tracker (Smart Eye Aurora) was positioned at the bottom of the monitor, with the distance between participants and eye tracker maintained between 60 and 70 cm. The eye tracker was calibrated using a 9-point calibration array followed by a 4-point validation array, and pupil size and eye-to-sensor distance were recorded at 60 Hz. The participants wore open-back headphones (Sennheiser HD600) and a microphone (C555L, AKG) was placed 5 cm from the mouth along the 45-degree axis: see Fig. 2

Figure 2. Experimental setup

The participants read the sentences in eight trials with four noise conditions and two speech production techniques. The order of the noise conditions and speaking styles was randomized.

2.3 Statistical Analyses

Self-Rating of Mental Demand: A one-way repeated-measures ANOVA evaluated the effect of the speaking condition on the self-rating of mental demand, with speaker as a within-subject factor. Post-hoc analysis with a pairwise t-test and Bonferroni correction identified significant differences in the ratings.

Pupillometry Data: The average pupil size between the right and left eyes was calculated and used for analysis. A three-way repeated-measures ANOVA evaluated the effects of speaking style, noise and order of the trial on pupil size. Post-hoc analysis with the Tukey HSD test determined significant differences in pupil size.

3. RESULTS

3.1 The effect of noise type and speech production techniques on self-rating of mental demand

The highest mental demand rating was reported in the MT noise condition, with an average of 11.4 (SE±2.94), whereas the lowest rating was reported for Quiet, with an average of 1.0 (SE±0.32). The average ratings for the RevMT noise condition was 7.4 (SE±1.44), and 4.0 (SE±1.41) for the SS noise. In clear speech, the highest mental demand rating was for the MT noise, with an average of 12.2 (SE±1.85), while the lowest rating was for Quiet, with an average of 1.6 (SE±0.75). The average rating for the RevMT noise was 8.4 (SE±2.27), and 3.4 (SE±1.03) for the SS noise: see Fig. 3.

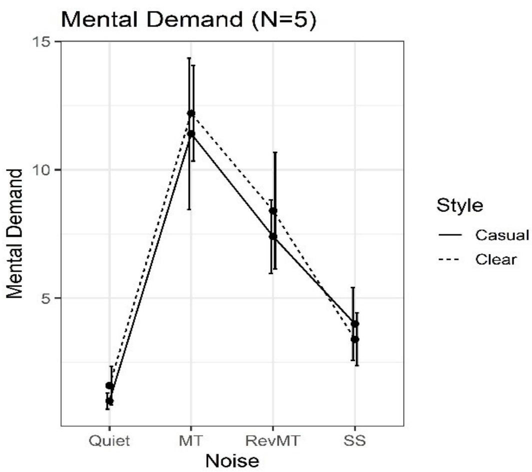

Figure 3. Average rating of mental demand for each noise condition. Error bars indicate standard error.

3.2 The effect of noise type and speech production techniques on pupil dilation

A total of 133,255 pupil measures were obtained across the participants. A repeated measures ANOVA was conducted to examine the effects of speaking style and noise type on average pupil size. Results indicated a significant main effect of speaking style, $F(1, 133236) = 46.47$, $p < .001$. There was also a significant main effect of noise type, $F(3, 133236) = 725.47$, $p < .001$. Furthermore, the analysis revealed a significant interaction between speaking style and noise type, $F(3, 133236) = 245.02$, $p < .001$. The main effect of order, which represents the order in which the stimuli were presented, was also found to be significant, $F(7, 133236) = 2164.69$, $p < .001$.

Post-hoc analysis using the Tukey HSD indicated a significant difference between the speaking styles, with clear speech showing a smaller pupil size than casual speech ($p_{adj} < 0.001$). For noise type, significant differences were found between all pairs of noise conditions, except between SS noise and RevMT noise ($p_{adj} < 0.85$).

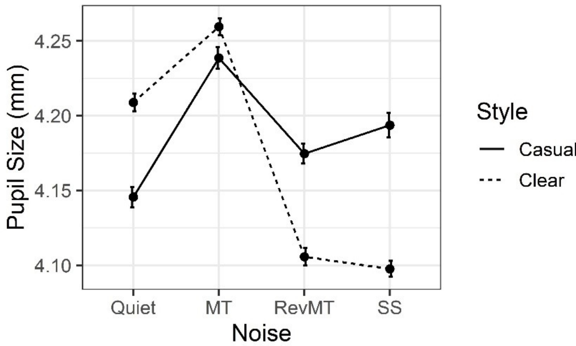

Figure 4. Line plot showing pupil size for all experimental conditions. Error bars indicate standard error.

Regarding the interaction between speaking style and noise type, a Tukey HSD post-hoc test was conducted to further examine the significant interaction between speaking style and noise type on average pupil size. While most pairwise comparisons demonstrated significant differences, several pairs did not yield significant differences in average pupil size. Specifically, no significant differences were found between Casual:RevMT and Clear:Quiet ($p = 0.056$), as well as between Casual:MT and Clear:MT ($p = 0.322$). Additionally, the pair Casual:SS and Casual:RevMT did not yield a significant difference ($p = 0.502$), and the difference between Clear:SS and Clear:RevMT was also not significant ($p = 0.980$): see Fig. 4.

3.3 The effect of trial order on pupil dilation

The Tukey HSD test revealed statistically significant differences in mean values between most order pairs ($p_{adj} < 0.001$), except for the comparison between orders 4 and 6 ($p_{adj} = .982$) and orders 7 and 8 ($p_{adj} = .112$): see Fig. 5.

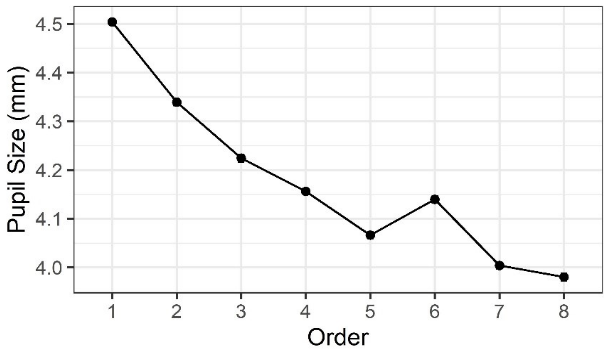

Figure 5. Line plot showing the average pupil size for all experimental trials. The error bars indicate standard error. Note that the small range of standard errors made the error bars invisible in this figure.

A Spearman's rank correlation was conducted to examine the relationship between average pupil size for each experimental condition and the mental demand ratings. The results showed a non-significant, weak positive correlation between the two variables (rs = 0.22, S = 8339.7, p = 0.18).

4. DISCUSSION

4.1 Noise type affects speakers' self-rating on mental effort

Clear speech has been endorsed as a therapeutic technique in earlier research, primarily due to the minimal training needed to enhance intelligibility [8, 13]. However, conscious effort is required when producing clear speech, particularly when instructed to do so. It was initially hypothesized that clear speech would demand a higher level of mental effort compared to casual speech. Contrary to this hypothesis, the findings revealed no significant increase in subjective mental effort ratings for clear speech, supporting the clinical rationale for its use instead.

With respect to noise type, we hypothesized that speech production in noisy environments would lead to increased cognitive load, with the most significant mental demand being in the MT noise condition due to its informative nature. The results lend partial support to this hypothesis, as the MT noise condition exhibited the highest mental demand rating, followed by the RevMT noise condition. The quiet and SS noise conditions showed the lowest mental demand, with no notable difference between them. This lack of difference could indicate that speakers were able to effectively adapt to these noise conditions. However, to confirm these results, a larger study is needed.

4.2 Speech style and noise type affect pupil measures

This study aimed to objectively measure the cognitive load associated with speech production styles and noise types, using pupillometry. We hypothesized that speaking in clear speech and noise would increase cognitive load, which is observed as a change in pupil size. These results partially support our hypotheses. The pupil size was significantly greater in MT noise than in quiet, RevMT noise, and SS noise for both casual and clear speech. There was no significant difference between casual and clear speech in MT noise, indicating that the effect of informational-masking noise has a greater impact on cognitive load than modifying speech to produce clear speech. The largest pupil size in MT noise suggests that the cognitive load was the highest in this type of noise, corroborating the subjective mental effort rating. As predicted, these findings suggest that noise with higher information content significantly increases the cognitive load during speech production, which is consistent with a previous report on speech perception in noise.

The data also illustrated a complex interaction between speech style and noise. Pupil size across noise conditions was smaller for clear speech than for casual speech, suggesting that producing clear speech requires less cognitive load than casual speech. An inspection of the data for each noise condition shows that the effect of the speaking style depends on the noise type. In quiet conditions, pupil size was greater for clear speech than for casual speech, suggesting that producing clear speech increases the cognitive load. This result indicates that clear speech may be easier to produce than other speech production techniques, yet still requires greater cognitive load than speaking in a habitual manner, especially in a quiet environment. The increase in the cognitive load for clear speech in quiet conditions was not indicated by the mental effort rating. A potential explanation for this discrepancy is that the mental effort associated with producing clear speech is subtle, which makes it difficult for speakers to detect.

Interestingly, the pupil size for clear speech was significantly smaller than that for casual speech in the energetic-masking noises. One possible explanation is that the noise naturally elicited the need for over-articulating speech sounds, thereby reducing the cognitive load required for intentionally adjusting speech. The greater pupil size for casual speech in these noises may imply that the speakers made some effort to resist the Lombard effect in an attempt to keep their speech habitual. If so, more explicit instruction and explanation for "casual" speech may have been needed. An alternative explanation is that the speakers found using clear speech under these noise conditions were impossible and disengaged from the task. Another important observation is that the pupil size for clear speech did not differ from that for casual speech produced in the RevMT and SS noises. This lack of difference may imply that using clear speech in a quiet environment requires as much cognitive effort as speaking in the presence of an energetic-masking noise.

The previous studies have demonstrated that pupil size can be a reliable measure of cognitive effort based on its significant correlation with mental effort ratings [21-22]. The mental effort rating did not correlate significantly with pupil size in our study. Caution is necessary when interpreting these results because of the significant order effect on pupil size. Despite randomization of experimental conditions, pupil size was the largest in the first trial and decreased in subsequent trials, which may be attributed to the initial arousal level of the speakers decreasing over time. Future studies should determine the number of trials needed to control for order of effects. Additionally, the lack of correlation between subjective and objective measures may also be due to the small sample size. Further studies with larger sample sizes and more trials are necessary to validate our results.

4.3 Limitations

Apart from small sample size, the current study faced multiple other limitations. Firstly, only one noise level was used for each noise type. Given that noise level can influence speech production, future research should examine the impact of varying noise levels on both speech production and cognitive effort. Furthermore, the study's participants were exclusively young adults with normal hearing capabilities and healthy voices. Additionally, the research employed a reading task, whereas the effect of noise type on speech tasks might differ for spontaneous speech. Prior research has shown that this effect was only present in spontaneous speech tasks [23] and not in reading-aloud tasks [24]. Therefore, assessing the applicability of these findings to other age groups, disordered populations, and real-life communication scenarios necessitates further exploration.

5. CONCLUSIONS

In summary, this study, although limited by a small sample size, demonstrates the feasibility of using pupillometry to examine the cognitive demands during speech production in various noise conditions. To the best of our knowledge, this study is the first investigation to employ pupillometry for evaluating cognitive load associated with speech modification and speaking in noise. The findings provide preliminary insights into the factors that influence communication in noisy environments, which could have implications for speech therapy techniques and the development of interventions for individuals with communication difficulties. While the results should be interpreted with caution due to the study's limitations, they highlight the value of further research utilizing pupillometry to enhance our understanding and ability to support individuals in effectively managing their communication challenges in diverse settings.

6. ACKNOWLEDGMENTS

This project was funded by the University of Illinois Campus Research Board Pilot Grant, awarded to Keiko Ishikawa.

7. REFERENCES

[1] J. Iwarsson, D. J. Morris, and L. W. Balling, "Cognitive Load in Voice Therapy Carry-Over Exercises," *J. Speech Lang. Hear Res.*, vol. 60, no. 1, pp. 1-12, 2017.

[2] Haith, A.M., and Krakauer, J.W.: 'The multiple effects of practice: skill, habit and reduced cognitive load', *Current opinion in behavioral sciences,* vol. 20, pp. 196-201, 2018.

[3] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Mal. de L'Oreille et du Larynx*, pp. 101-119, 1911.

[4] M. Basner et al., "ICBEN review of research on the biological effects of noise 2011-2014," Noise Health, vol. 17, no. 75, pp. 57, 2015.

[5] D. C. Button, D. G. Behm, M. Holmes, and S. N. Mackinnon, "Noise and muscle contraction affecting vigilance task performance," Occup. Ergon., vol. 4, no. 3, pp. 157-171, 2004.

[6] R. W. Hughes, "Auditory distraction: A duplex‐mechanism account," PsyCh Journal, vol. 3, no. 1, pp. 30-41, 2014.

[7] G. Kidd Jr, C. R. Mason, V. M. Richards, F. J. Gallun and N. I. Durlach, "Informational masking. Auditory perception of sound sources," in *Auditory Perception of Sound Sources*, pp. 143-189, 2008.

[8] J. Rennies, V. Best, E. Roverud and G. Kidd Jr, "Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort," in *Trends in Hearing*, vol. 23, pp. 1-21, 2019.

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

[10] A. R. Bradlow, N. Kraus, and E. Hayes, "Speaking clearly for children with learning disabilities," *J. Speech Lang. Hear. Res*., 2003.

[11] M. A. Picheny, N. I. Durlach, and L. D. Braida, "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Lang. Hear. Res*., vol. 29, no. 4, pp. 434-446, 1986.

[12] R. Smiljanić and A. R. Bradlow, "Temporal organization of English clear and conversational speech," *J. Acoust. Soc. Am*., vol. 124, no. 5, pp. 3171-3182, 2008.

[13] R. Smiljanić and A. R. Bradlow, "Speaking and hearing clearly: Talker and listener factors in speaking style changes," *Lang. Linguist. Compass*, vol. 3, no. 1, pp. 236-264, 2009.

[14] J. Gartner-Schmidt, S. Gherson, E. R. Hapner, J. Muckala, D. Roth, S. Schneider, and A. I. Gillespie, "The development of conversation training therapy: A concept paper," *J. Voice*, vol. 30, no. 5, pp. 563-573, 2016.

[15] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," Educational Psychologist, vol. 38, no. 1, pp. 63-71, 2003.

[16] S. Sirois and J. Brisson, "Pupillometry," in Wiley Interdisciplinary Reviews: *Cognitive Science,* vol. 5, no. 6, pp. 679-692, 2014.

[17] S. Mathôt, "Pupillometry: Psychology, physiology, and function," in *Journal of Cognition,* vol. 1, no. 1, 2018.

[18] M. B. Winn, D. Wendt, T. Koelewijn and S. E. Kuchinsky, "Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started," in Trends in Hearing, vol. 22, pp. 1-32, 2018.

[19] M. Nilsson, S. D. Soli and J. A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," in *The Journal of the Acoustical Society of America,* vol. 95, no. 2, pp. 1085-1099, 1994.

[20] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology,* vol. 52, pp. 139-183, North-Holland, 1988.

[21] S. Alhanbali, K. J. Munro, P. Dawes, P. J. Carolan and R. E. Millman, "Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy," in *International Journal of Audiology*, vol. 60, no. 10, pp. 762-772, 2021.

[22] H. Devos, K. Gustafson, P. Ahmadnezhad, K. Liao, J. D. Mahnken, W. M. Brooks and J. M. Burns, "Psychometric properties of NASA-TLX and index of cognitive activity as measures of cognitive workload in older adults," in *Brain Sciences,* vol. 10, no. 12, pp. 994, 2020.

[23] M. Cooke and Y. Lu, "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," in *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 2059-2069, 2010.

[24] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise," in *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 3261-3275, 2008.