scientific reports

OPEN

Cognitive load associated with speaking clearly in reverberant rooms

Keiko Ishikawa[1✉], Silvia Murgia[2], Hannah Li[2], Elisabeth Renkert[2] & Pasquale Bottalico[2]

Communication is a fundamental aspect of human interaction, yet many individuals must speak in less-than-ideal acoustic environments daily. Adapting their speech to ensure intelligibility in these varied settings can impose a significant cognitive burden. Understanding this burden on talkers has significant implications for the design of public spaces and workplace environments, as well as speaker training programs. The aim of this study was to examine how room acoustics and speaking style affect cognitive load through self-rating of mental demand and pupillometry. Nineteen adult native speakers of American English were instructed to read sentences in both casual and clear speech—a technique known to enhance intelligibility—across three levels of reverberation (0.05 s, 1.2 s, and 1.83 s at 500–1000 Hz). Our findings revealed that speaking style consistently affects the cognitive load on talkers more than room acoustics across the tested reverberation range. Specifically, pupillometry data suggested that speaking in clear speech elevates the cognitive load comparably to speaking in a room with long reverberation, challenging the conventional view of clear speech as an 'easy' strategy for improving intelligibility. These results underscore the importance of accounting for talkers' cognitive load when optimizing room acoustics and developing speech production training.

From whispering in a carpeted library to speaking in the echoing expanse of a cathedral, people communicate in a variety of room environments. Some of these environments may present challenges to speaking intelligibly due to the smearing effect caused by reverberation. To navigate these adverse conditions, talkers monitor their speech quality and subsequently adjust the way they produce it. This speech modification is an excellent illustration of Lindblom's Hyper and Hypo-articulation (H&H) model, which posits that talkers modify their speech in response to various communicative demands and constraints, aiming to balance communication efficiency and effort[1]. In situations that present communication barriers, such as a talker having a speech or voice disorder or a communication partner having hearing loss, the necessity to hyperarticulate becomes even more pronounced. The adjustment of the speech production mechanism and monitoring of its output likely require an increased allocation of cognitive resources and attention. While the cognitive load associated with room acoustics has been well-documented in the context of speech perception[2–4], its association with speech production has not been studied as thoroughly. Understanding the cognitive load associated with speech modification and room acoustics would provide valuable insights for speech training programs and architectural design.

Effect of cognitive load on speech production

"Cognitive load" refers to the mental resources required to process information and perform a specific task[5]. The dynamic relationship between cognitive load and speech production is well-documented. For example, Lively et al. investigated the effect of cognitive load on speech production through talkers engaging in a compensatory visual tracking task[6]. They observed changes in speech characteristics, including increased amplitude and variability, decreased spectral tilt, and altered speaking rate. Lively et al. further examined the acoustic adaptations during the visual tracking task for their intelligibility and reported that certain modifications, particularly in amplitude, could enhance speech intelligibility. These intelligibility-enhancing changes led the authors to conclude that their findings are in line with the premise of the H&H model. Similarly, MacPherson showed that cognitive load, manipulated through a Stroop task in congruent and incongruent conditions, affects variability in speech motor movements[7]. Furthermore, it has been reported that significant interference can arise from concurrent tasks regulated by brain networks in close anatomical proximity[8,9], suggesting that impact of cognitive load on speech production is affected by the nature of secondary tasks. While these studies illustrate

[1]Department of Communication Science and Disorders, University of Kentucky, Lexington, Kentucky, USA. [2]Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA. ✉email: ishikawa@uky.edu

how cognitive load influences speech production, the reverse—how speech production may impact cognitive load—remains underexplored, underscoring a unidirectional understanding in our current body of knowledge.

Clear speech

The most well-known speech adaptation for challenging acoustic environments is the Lombard effect. Examples of such challenging environments include classrooms, open-plan offices, and transportation hubs like train stations, characterized by large, echoing spaces and constant ambient noise. This effect typically occurs when talkers increase vocal effort in response to environmental acoustics that mask speech intelligibility. The resultant changes in speech acoustics include changes in speech rate and spectral properties[10,11]. While it is commonly associated with noisy environments, reverberation also prompts speech adjustments, eliciting various acoustic changes including an increase in the duration of voicing and silent periods[12]. While this adaptation occurs automatically, talkers can also intentionally change the way they speak in an attempt to enhance the intelligibility of their speech. One common strategy, particularly in challenging listening environments, is clear speech[13,14]. The intelligibility benefit brought by clear speech has proven to be valuable in various contexts, such as for hearing-impaired listeners[15–19] and second language learners[13,14]. The use of clear speech elicits various changes in speech acoustics. Global changes include increased intensity and pitch range[20]. Temporal adjustments include a decreased speech rate, extended vowel duration, more frequent release of stop consonants and word-final consonants, and increased plosive duration[15,20–22].

Effect of room acoustics on speech perception and production

It has been well-demonstrated that room acoustics, especially reverberation, affect speech perception. Increased reverberation and noise levels negatively impact speech intelligibility, requiring greater listening effort, which is the cognitive load listeners allocate to understand the speech[2–4]. Reverberation also affects speech production. Bottalico, Graetzer, and Hunter explored how vocal effort is affected by speech style, room acoustics, and short-term vocal fatigue[23]. They recorded 20 talkers reading aloud in various settings—anechoic, semi-reverberant, and reverberant rooms—amid background classroom babble noise. Their findings showed an increase in sound pressure level (SPL) and perceived effort during loud reading, which diminished with the introduction of reflective panels and in environments with longer reverberation times.

Capturing listening effort has been a key interest in speech perception research, with pupillometry extensively used to measure cognitive load allocated by listeners in understanding speech[24,25]. This technique is regarded as a psychophysiological measure of cognitive load, based on the observation that the pupil's size changes in response to various cognitive demands. Specifically, pupil dilation tends to increase as cognitive load rises. In contrast, research that examined mental effort, or cognitive load, involved in speaking under various acoustic condition relied on self-reporting[23,26]. For instance, Ishikawa, Li, and Coster explored the mental effort involved in speaking casually versus clearly amid background noise[26]. They found that talkers reported a higher mental demand in multi-talker noise compared to reversed multi-talker noise and speech-shaped noise, but they did not report any difference between casual and clear speech. This absence of difference might suggest a limitation of self-reporting in capturing subtle changes in mental effort required for speech modification. Alternatively, it could imply that using clear speech does not significantly increase mental load for the talkers.

Previous research has documented the effect of acoustic environments on speech production, as well as cognitive load related to speech perception in various acoustic environments. Understanding the effect of room acoustics and speech modification on cognitive load is crucial in optimizing communication environments and mitigating the effects of challenging room acoustics. Furthermore, this understanding will provide a critical foundation for designing speaker training programs that are effective yet cognitively manageable, ensuring that individuals can improve their communicative abilities without being burdened by the complexity of the training. Despite its significance, little is known about the cognitive load associated with speech modification and room acoustics. Rooted in the H&H model, this study examines the changes in cognitive load in response to different speaking styles and reverberation levels. It is hypothesized that clear speech, which requires intentional changes to enhance intelligibility, increases cognitive load compared to casual speech. Additionally, it is hypothesized that longer reverberation times will also increase cognitive load. Self-reports and pupillometry are used to measure the change in cognitive load. Lastly, the talkers' ability to use clear speech in given acoustic conditions is evaluated.

Methods

Participants

Nineteen adult native speakers of American English (age range: 18–35 years) with no history of speech, language, or hearing disorders participated in this study. All participants provided written informed consent prior to participation and were compensated for their time. The experimental protocols for this study were approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign (#19215). All experiments were performed in accordance with relevant guidelines and regulations.

Experimental design

This study used a within-subjects design with speech production style (casual vs. clear) and room acoustics (No Effect, Small Room and Large Room) as the independent variables.

Acoustic simulation procedure

The virtual acoustic environments were simulated using a real-time effect processor (MX400, Lexicon). Acoustics of the virtual environments were characterized with oral-binaural impulse responses (IRs) calculated with the convolution method. A Class 1 microphone (M2211, NTi Audio) was calibrated with a Class 1 Sound Calibrator

(NTi Audio) and then placed at a fixed distance of 15 cm from the corner of the mouth speaker of a Head and Torso Simulator (HATS) (45BC KEMAR HATS, GRAS) at a 45-degree angle. The choice of a 15 cm mouth-to-microphone distance in the current study is based on several critical considerations. First, this distance helps mitigate the proximity effect, which can exaggerate low-frequency responses when the microphone was too close to the speaker, thereby ensuring a more accurate representation of the speaker's voice across various frequencies. Additionally, a 15 cm distance reduces sensitivity to speaker movements, maintaining consistent acoustic measurements even if the speaker shifts slightly. Furthermore, this distance minimizes the delay between the uttered voice and its transmission through the processing loop and back to the participant's headphones, keeping the delay smaller than the just noticeable difference for delay perception[27].

Room acoustic T30 conditions (ISO 3382-2, 2008)[28] of the virtual scenarios were obtained from impulse responses (IRs) calculated with the convolution method. An exponential sweep signal (from 100 to 10,000 Hz) was emitted from the mouth speaker of the HATS and captured by the microphone. The captured signal was processed in real-time to add reverberation using the real-time effect processor. The processed signal was sent to the "ears" of the HATS via open-back headphones (HD600, Sennheiser) and the recorded sweep was deconvolved with the emitted sweep inverted on the time axes to obtain the IR, as described by Pelegrín-García and Brunskog[29]. The average T30 measurements for combined 500 Hz and 1 k Hz octave bands were determined for the single-wall sound-proof booth and the two simulated environments. The average T30s were 0.05 s, 1.2 s, and 1.83 s at 500–1000 Hz for the sound-proof booth (No Effect), short reverberation (Small Room) and long reverberation (Large Room) conditions, respectively. The delay between the real-time signal and processed signal fed back to the headphones was less than 5 ms, which was below the threshold of noticeable difference[30]. The average T30 of the sound-proof booth represents an anechoic environment without acoustic reflections. The average T30 of the Small Room is an excellent simulation of a very small room used for dialog and voiceover applications. The average T30 of the Large Room emulates the acoustics of a concert hall—a space large enough to contain an orchestra and an audience[31].

Data collection procedures

After consent was obtained, the participants underwent a brief training to ensure that they were able to differentially produce casual and clear speech. The training was provided by two graduate students in speech-language pathology and lasted less than 10 min. The participants were instructed to speak naturally for casual speech. The instruction for clear speech was as follows: "Please speak as if you were leaving an important message on someone's phone. Clearly articulate your speech sounds so that they can understand you." The participants were deemed ready for the experiment when they could switch between casual and clear speech with 100% accuracy. After confirming that participants were able to produce both speech types distinctly, they were taken to a single-wall sound-proof booth, in which the experiment was conducted. The participants were seated in front of a 27-inch computer monitor, which displayed stimuli presentation slides. The stimuli consisted of a set of three slides for each reverberation and speaking style condition, for a total of eighteen sets (3 room conditions × 2 speaking styles × 3 repetitions). The first slide displayed a baseline cross at the center of the screen for 5 s. The second slide displayed two lists of 10 sentences from the Hearing in Noise Test (HINT)[32] and Speech in Noise (SPIN) test[33]. The last slide of the set displayed a 20-point scale for rating the degree of the participant's mental demand, modeled after the NASA Task Load Index scale[34]. (Fig. 1) Participants were instructed to read sentences in both casual and clear speech under each of the three reverberation conditions. The order of conditions was counterbalanced across participants to minimize order effects.

During the experiment, the participant's speech was recorded with a Class 1 microphone (M2211, NTi Audio) placed at a fixed distance of 15 cm from the corner of the participant's mouth at a 45-degree angle. The microphone was calibrated prior to the experiment. The captured audio signal was split into two channels. The first channel was used for direct recording of the participant's speech. This channel contained the anechoic
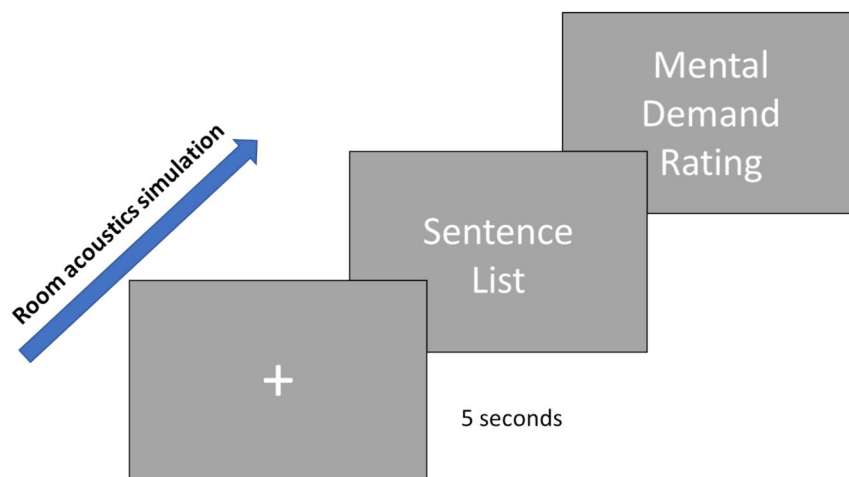
Fig. 1. Schematic description of a trial.

speech, used for the following analyses. The speech signal was sent to an external soundboard (UH-7000, TAS-CAM, Teac Corporation), which was connected to a laptop computer (Latitude 7480, Dell). The signal was then recorded using Audacity 3.0.0 (SourceForge) at the sampling rate of 44,100 Hz and bit depth of 16. The second channel was used to send the signal to the real-time effect processor, and the processed signal was then sent to the open-back headphones, through which the participants heard their speech with reverberation. The open-back headphones were used so that participants could hear their direct sound unaltered and the reverberant tail through the headphones.

Pupillary changes were captured by an eye tracker (Aurora, Smart Eye), which was positioned at the bottom of the computer monitor. As specified by the manufacturer of the eye tracker and data collection software (iMotions), the distance between the participants and the eye tracker was maintained between 60 and 70 cm. The participants were instructed to minimize their head movement during the experiment. The eye tracker calibration was systematically conducted at fixed intervals, using a predetermined 9-point calibration array followed by a 4-point validation array. Pupil size and eye-to-sensor distance were then continuously recorded at 60 Hz during the experiment. To prevent any pupillary reactions to changes in ambient light, consistent brightness was maintained throughout the experiment.

Pupillometry data preprocessing

Pupil measurements during blinks, as indicated by the iMotion software, were removed from the dataset. Following the recommendation by Hershman, Milshtein, and Henick[35], the pupillometry data were preprocessed to obtain normalized, task-related pupil size changes using the divisive baseline correction method. For each participant and task (i.e., reading sentences aloud in casual or clear speech), the average pupil size during the 5-s display of the fixation baseline cross was computed. Subsequently, the average task-related pupil size was calculated for the entire duration of each task, including short time frames when no speech was produced between sentences. The choice to include these time frames in the analysis was made to mirror how people typically read aloud sentences, thereby enhancing the ecological validity of our findings. This approach also allows for variability in the duration it takes for participants to complete the task, similar to the method used by Hess and Polt[35,36], which reported changes in pupil size during mathematical problem-solving. The average task-related pupil size was then divided by the average baseline pupil size for each trial, resulting in a ratio that accounts for individual differences in baseline pupil size.

Acoustic analysis

Participants' speech production behaviors were acoustically assessed via speech rate (syllables per second) and intensity. Prior to the analyses, trained research staff manually edited the recordings to extract only the speech portions, using a spectrogram to accurately identify the start and end of each speech segment. This process involved visually and auditorily inspecting the spectrogram for specific acoustic cues such as sudden increases in energy or distinct frequency patterns that signify the commencement and cessation of speech. These segments were then edited out and saved in .wav format. Lastly, we measured the rate and intensity of each of the 20 sentences in a trial using Praat[37] and its associated scripts[38].

Statistical analyses

A repeated measures ANOVA was conducted to examine the effects of speech production style and room acoustics on subjective rating of mental demand, pupillometry, and acoustic measurements. Post-hoc comparisons were conducted using Bonferroni-adjusted pairwise comparisons. Significance levels were set at $p < 0.05$.

Ethics declarations

The experimental protocols for this study were approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign (#19215).

Results
Subjective rating of mental demand

The results of the repeated measures ANOVA revealed significant main effects for both room acoustics ($F(2, 320) = 4.79$, $p = 0.009$, $\eta^2 = 0.029$) and speech production style ($F(1, 320) = 48.93$, $p < < 0.001$, $\eta^2 = 0.133$), indicating that the participants' self-ratings were influenced by the room acoustics and the speech production style they employed.

The post-hoc analysis using pairwise t-tests with Bonferroni correction revealed significant differences in the ratings between casual and clear speech styles across all room conditions, with No Effect room, $t(56) = -4.654$, $p_{adj} < < 0.001$; Small Room, $t(56) = -4.187$, $p_{adj} = 0.0001$; and Large Room, $t(56) = -4.018$, $p_{adj} = 0.0002$. In terms of the difference between room conditions, significant differences were found only between No Effect and Small Room, $t(56) = -3.394$, $p_{adj} = 0.004$ in the casual speech style. No significant differences were found between No Effect and Large Room, $t(56) = -2.143$, $p_{adj} = 0.109$, or between Small Room and Large Room, $t(56) = 0.553$, $p_{adj} = 1$. In the clear speech style, no significant differences were observed between any of the room conditions, with $t(56) = -1.555$, $p_{adj} = 0.378$ for No Effect vs. Small Room; $t(56) = -2.261$, $p_{adj} = 0.083$ for No Effect vs. Large Room; and $t(56) = -0.327$, $p_{adj} = 1$ for Small Room vs. Large Room. (Fig. 2).

Pupillometry

The results of the repeated measures ANOVA showed no significant effect of room acoustics on the ratio of pupillary change ($F(2, 320) = 1.225$, $p = 0.295$, partial $\eta^2 = 0.008$). However, there was a significant effect of
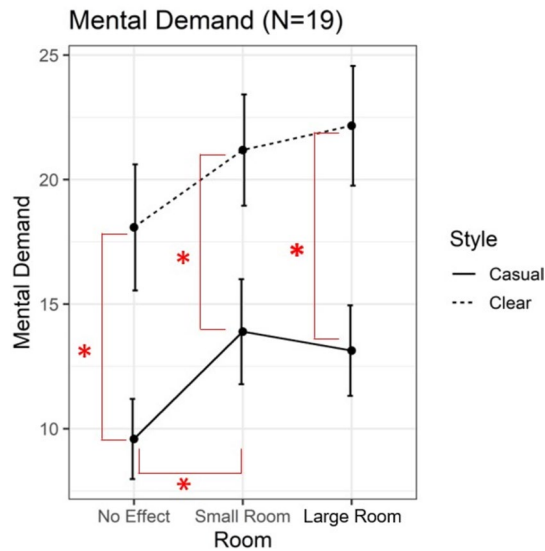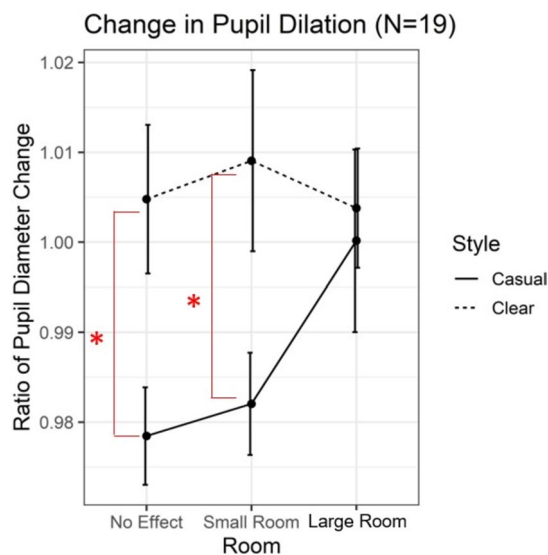
Fig. 2. Mental demand ratings for different room acoustic conditions and speaking styles. Error bars indicate standard error. Asterisks indicate a pair of conditions that yielded a statistically significant difference.

speech production style on the pupillary response ($F(1, 320) = 12.131$, $p < 0.001$, partial $\eta^2 = 0.04$). The interaction between room acoustics and speech production was not statistically significant ($F(2, 320) = 1.988$, $p < 0.139$, partial $\eta^2 = 0.01$).

The results of the pairwise t-tests with Bonferroni correction for post-hoc analysis revealed significant differences in the pupillary response between casual and clear speech production styles in both No Effect ($t(56) = -2.932$, $p_{adj} = 0.005$) and Small Room ($t(56) = -3.021$, $p_{adj} = 0.004$). There was no significant difference between casual and clear speech production styles in Large Room ($t(56) = -0.338$, $p_{adj} = 0.736$) (Fig. 3).

Speech rate

Due to equipment failure, the audio recordings from three participants were not captured. Consequently, the analyses were conducted on data from the remaining sixteen participants. The results of the repeated measures ANOVA revealed significant main effects for speech production style ($F(1, 284) = 35.67$, $p < < 0.001$, $\eta^2 = 0.460$), indicating that the speech rate of the clear speech style was significantly slower than that of the casual speech style. However, there was no a significant effect of room acoustics on the speech rate ($F(2, 284) = 0.16$, $p < 0.583$, $\eta^2 = 0.002$).

Fig. 3. Ratio of pupil diameter change for different room acoustic conditions and speaking styles. Error bars indicate standard error. Asterisks indicate a pair of conditions that yielded a statistically significant difference.

The results of the pairwise t-tests with Bonferroni correction for post-hoc analysis revealed significant differences in the speech rate between casual and clear speech production styles in both the No Effect ($t(48) = 13.6$, $p_{adj} < < 0.001$), Small Room ($t(48) = 14.3$, $p_{adj} < < 0.001$) and Large Room ($t(48) = 13.2$, $p_{adj} < < 0.001$) (Fig. 4).

Speech Intensity

The results of the repeated measures ANOVA revealed significant main effects for both room acoustics ($F(2, 269) = 3.81$, $p = 0.023$, $\eta^2 = 0.028$) and speech production style ($F(1, 269) = 195.49$, $p < 0.001$, $\eta^2 = 0.421$).

The results of the pairwise t-tests with Bonferroni correction for post-hoc analysis revealed that the intensity was significantly lower in Small Room compared to No Effect, $t(47) = 2.71$, $p_{adj} = 0.028$ for the casual speech style. There were no significant differences between the No Effect and Large Room, $t(47) = 0.483$, $p_{adj} = 1$, and between Small Room and Large Room, $t(47) = -2.02$, $p_{adj} = 0.149$. For the clear speech style, the intensity was significantly lower in the Small Room condition compared to the Large Room, $t(47) = -2.80$, $p_{adj} = 0.022$. There were no significant differences in intensity between the No Effect and Small Room conditions, $t(47) = 1.95$, $p_{adj} = 0.172$, and between the No Effect and Large Room conditions, $t(47) = -0.237$, $p_{adj} = 1$.

The tests also revealed that the intensity of the casual speech style was significantly lower than that of the clear speech style across all tested room conditions (No Effect, $t(47) = -9.15$, $p_{adj} < < 0.001$; Small Room, $t(47) = -7.30$, $p_{adj} < < 0.001$; Large Room, $t(47) = -7.54$, $p_{adj} < < 0.001$) (Fig. 5).

Discussion

The fundamental goal of communication is to gain mutual understanding between conversational partners. Achieving this goal becomes more challenging in certain environments, such as in reverberant and noisy rooms, or in situations with communication barriers. Under these conditions, speakers must adjust their speech production and monitor its output to ensure intelligibility is maintained. These adjustments and monitoring processes likely increase the allocation of cognitive resources, yet the extent of resources required for these processes remains poorly understood. To address this knowledge gap, this study examined how speech modification and room acoustics affect a talker's cognitive load.

The results support the hypothesis that cognitive load is elevated when speaking in clear speech compared to casual speech, as evidenced by both the subjective ratings of mental demand and the pupillometry data. Clear speech consistently led to higher ratings of mental demand across all room conditions. Furthermore, pupillometry data revealed that the ratio of pupil diameter change was significantly greater for clear speech than for casual speech in the No Effect and Small Room conditions, suggesting an increased cognitive load for clear speech. For casual speech, the ratio of pupil diameter change was less than 1, indicating that pupils constricted more compared to the baseline. This constriction is believed to be an artifact of the stimuli presentation slide; unlike the baseline condition, which displayed only a single cross in the middle, the speaking trials showed twenty sentences, making the sentence slides brighter and leading to pupil constriction. Despite this, it is important to emphasize that pupils dilated more for clear speech than for casual speech in these conditions. Additionally, the order of room conditions and speaking styles was randomized to control for the effect of slide content, supporting the conclusion that the observed pupillary responses primarily reflect the cognitive load differences between clear and casual speech styles.

The elevated cognitive load observed during clear speech production challenges the longstanding perception of clear speech being an "easy" method for enhancing intelligibility, traditionally thought to require minimal training for effective implementation[13,15]. Notably, the lack of significant differences in pupil response between
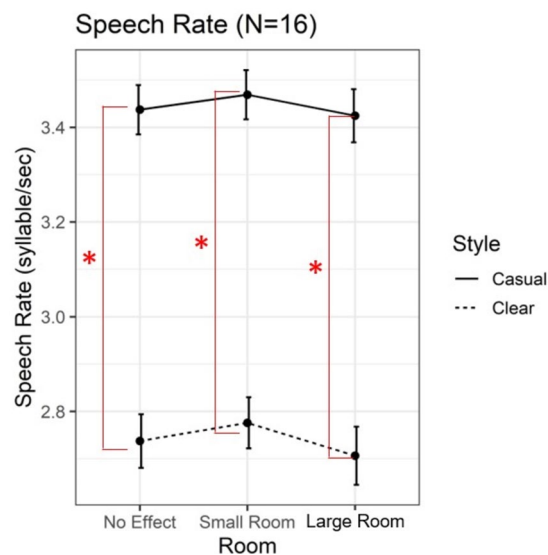
Fig. 4. Speech rate for different room acoustic conditions and speaking styles. Error bars indicate standard error. Asterisks indicate a pair of conditions that yielded a statistically significant difference.
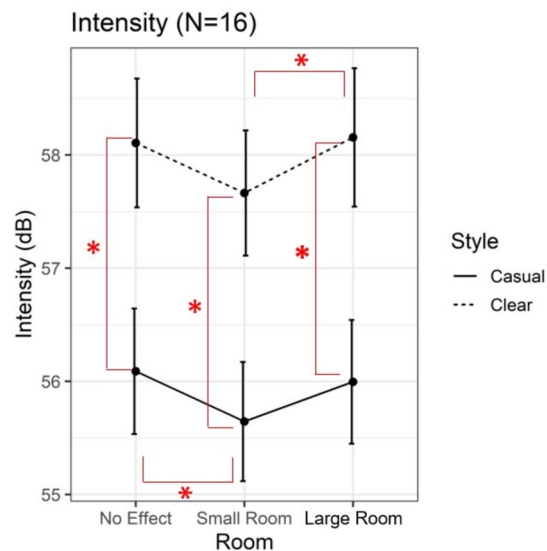
Fig. 5. Speech intensity for different room acoustic conditions and speaking styles. Error bars indicate standard error. Asterisks indicate a pair of conditions that yielded a statistically significant difference.

casual and clear speech in environments with extended reverberation times suggests that the cognitive effort required for clear speech mirrors that of casual speech under such conditions. This finding urges a deeper exploration into the cognitive demands of various speech modification techniques used in voice and speech therapy, potentially revealing even greater cognitive challenges. According to Cognitive Load Theory, cognitive overload can critically hinder learning capabilities[39]. Techniques that impose excessive cognitive demands may not only be difficult for patients to learn but also to apply in real-world scenarios, where multitasking is often necessary. Thus, by optimizing cognitive loads, therapists can facilitate a more effective learning process, increasing the likelihood that patients will successfully integrate and utilize new communication strategies in their daily interactions.

The expectation that reverberation would increase cognitive demand due to difficulty in monitoring speech appears intuitive, as reverberation can blur speech sounds, making it harder for speakers to hear their own speech accurately and adjust it in real time. However, the results provide minimal support for the hypothesis that longer reverberation times increase cognitive load. Subjective ratings indicated that room acoustics influenced mental demand, with significant differences found between the No Effect and Small Room conditions for the casual speech style. However, no change in mental demand was reported for clear speech across any room conditions. Moreover, the pupillometry data did not show a significant effect of room acoustics on pupillary responses for either casual or clear speech styles. The absence of the effect of room acoustics on cognitive load does not align with the premises of the H&H model that speakers engage in hyper-articulation as a strategic response to optimize communication. This unexpected result may suggest that conditions with longer reverberation times may have been needed to reveal its effect. To support this assumption, the acoustic examination of speech recordings indicated that the effects of room acoustics on speech production behaviors was minimal.

Contrary to our observations, previous literature has demonstrated the influence of acoustic environments on speech and voice production. For instance, Hodoshima, Arai, and Kurisu discovered that speech produced in reverberant conditions was more intelligible than speech in quiet settings[40]. Similarly, research on singers by Bottalico, Łastowiecka, Glasner, & Redman demonstrated that room acoustics significantly influence vibrato rate, extent, and pitch inaccuracy[41], indicating that singers modify their vocal production in response to different performance spaces. This adaptability might contrast sharply with non-singers or individuals with speech disorders, who may not inherently adjust their voice based on acoustic feedback or may lack the training to do so effectively. A study by Hodoshima, Arai, and Kurisu reported the greater intelligibility of speech produced in reverberation than in quiet with non-singers. They utilized reverberation times of 3.6 and 2.6 seconds[40]—longer than those used in our study, pointing towards the potential benefits of exploring more varied and realistic acoustic settings in future research. Extending investigations to include real-life therapy rooms, ranging from smaller individual to larger group spaces, could enrich our understanding of how different individuals, including those with cognitive impairments or neurological voice and speech disorders, adapt their vocal production. Such studies would not only validate the need for adaptive strategies but also potentially guide therapeutic interventions to manage cognitive load and improve communicative effectiveness across diverse acoustic environments.

To the best of our knowledge, this study is the first to utilize pupillometry to examine the cognitive load associated with speech modification in varying room acoustics. Employing self-reports and pupillometry as dual measures of cognitive load enables us to capture a more comprehensive picture of the mental effort involved in speech production. This approach is widely utilized in cognitive science because it reveals cognitive loads beyond what individuals can detect themselves, offering insights into unconscious cognitive processes[42]. The discrepancy between the subjective ratings and pupillometry data observed may suggest that the cognitive load associated with adjusting to room acoustics is subtle and might be overshadowed by the more pronounced effect

of speech production style. Other psychophysiological methods might be more sensitive to the nuanced effects of room acoustics on cognitive load. Alternatively, this discrepancy might indicate that the talker's perception of increased effort does not directly translate to a measurable physiological response in terms of pupillary change. The discrepancy is consistent with the broader literature on listening effort, indicating that different measures of listening effort (behavioral, subjective, and physiological) do not always correlate strongly with each other, suggesting that these measures may tap into distinct aspects of cognitive processing[43].

The observation that cognitive demand for producing clear speech remains constant across varying room acoustic conditions, including environments with long reverberation times, stimulates further inquiry into how speakers manage speech production in challenging acoustic environments. Two potential theories can be offered for the underlying mechanisms. The first, the invariance of cognitive load, suggests that the cognitive effort involved in producing clear speech is stable across different acoustic environments. This theory posits that engaging in clear speech production sets a fixed cognitive load that remains unaffected by changes in room acoustics. The second theory, cognitive prioritization for speech modification, posits that focusing on clear speech minimizes the impact of room acoustics on cognitive demand. While room acoustics might usually influence cognitive load, this theory argues that the deliberate focus on clear speech production can make these acoustic challenges secondary, highlighting a strategic redirection of cognitive resources towards speech clarity over environmental adaptation. Elucidating the underlying mechanisms is crucial for advancing speaker training methods. For instance, training programs could be tailored to either manage cognitive load during speech production or assist individuals in adapting to diverse acoustic environments, depending on their specific needs.

Limitations

This study, while providing valuable insights into the cognitive demands of speech production in various acoustic environments, has several limitations that warrant consideration. Firstly, the small sample size may limit the generalizability of our findings. Secondly, the reverberation times used in our experimental setups may not have been long enough to fully capture the impact of room acoustics on cognitive load. Additionally, the study's reliance on pupillometry and self-reports as the sole measures of cognitive load may not encompass all aspects of cognitive effort involved in speech production. While pupillometry provides a valuable objective measure, incorporating other psychophysiological markers could offer a more nuanced understanding of the cognitive processes at play.

Another potential limitation is the practice effect associated with the training protocol used to prepare participants for producing casual versus clear speech. Participants underwent several practice trials before recording the target stimuli, which might have enabled them to become accustomed to speaking clearly. This practice could have reduced the cognitive load required to produce clear speech during the experiment. Since participants' readiness was assessed solely based on their performance, it remains unclear to what extent practice altered the cognitive load. Consequently, the observed cognitive load during recordings might not fully reflect the actual cognitive effort involved in speech modification.

Finally, the study did not account for individual differences in speech production habits, auditory feedback sensitivity, or previous training in speech modification techniques, all of which could influence how speakers adjust to varying acoustic conditions. Acknowledging these limitations, our findings lay the groundwork for further research aimed at exploring the intricate relationship among cognitive load, speech production, and room acoustics, ultimately guiding the development of more effective communication strategies and therapeutic interventions.

Conclusion

Findings of this study underscore the importance of understanding the cognitive mechanisms underlying speech production and modification in various acoustic environments. Our findings illuminate the complexities of efforts underlying speech modification, challenging assumptions about the cognitive ease of producing clear speech. Specifically, we demonstrated that clear speech imposes a significant cognitive load on talkers, a load that does not markedly fluctuate with changes in room acoustics. This constancy suggests a potential invariance in cognitive load or a cognitive prioritization that renders acoustic challenges secondary when clarity in speech production is the focus.

Data availability

The data used in this study are not publicly available due to the sensitive nature of the participant information. Upon request, the corresponding author can provide de-identified data to qualified researchers for the purpose of replication and further analysis.

References

1. Lindblom, B. *Speech production and speech modelling* 403–439 (Springer, 1990).
2. Peng, Z. E. & Wang, L. M. Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. *J. Speech Lang. Hear. Res.* **62**, 1068–1081 (2019).
3. Prodi, N. & Visentin, C. A slight increase in reverberation time in the classroom affects performance and behavioral listening effort. *Ear Hear.* **43**, 460–476 (2022).
4. Rennies, J., Schepker, H., Holube, I. & Kollmeier, B. Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *J. Acoust. Soc. Am.* **136**, 2642–2653 (2014).
5. Paas, F. G. & Van Merriënboer, J. J. Instructional control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* **6**, 351–371 (1994).

6. Lively, S. E., Pisoni, D. B., Van Summers, W. & Bernacki, R. H. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *J. Acoust. Soc. Am.* **93**, 2962–2973 (1993).
7. MacPherson, M. K. Cognitive load affects speech motor performance differently in older and younger adults. *J. Speech Lang. Hear. Res.* **62**, 1258–1277 (2019).
8. Dromey, C. & Benson, A. Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. (2003).
9. Dromey, C. & Shim, E. The effects of divided attention on speech motor, verbal fluency, and manual task performance. (2008).
10. Garnier, M., Henrich, N. & Dubois, D. Influence of sound immersion and communicative interaction on the Lombard effect. (2010).
11. Lombard, E. L. signe de l'élévation de la voix (translated from French). *Ann. des Mal. l'oreille du larynx* **37**, 101–119 (1911).
12. Astolfi, A., Carullo, A., Pavese, L. & Puglisi, G. E. Duration of voicing and silence periods of continuous speech in different acoustic environments. *J. Acoust. Soc. Am.* **137**, 565–579 (2015).
13. Bradlow, A. R. & Bent, T. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* **112**, 272–284 (2002).
14. Smiljanić, R. & Bradlow, A. R. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Lang. Linguist. Compass* **3**, 236–264 (2009).
15. Ferguson, S. H. & Kewley-Port, D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* **112**, 259–271 (2002).
16. Payton, K. L., Uchanski, R. M. & Braida, L. D. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J. Acoust. Soc. Am.* **95**, 1581–1592 (1994).
17. Picheny, M. A., Durlach, N. I. & Braida, L. D. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Lang. Hear. Res.* **28**, 96–103 (1985).
18. Schum, D. J. Intelligibility of clear and conversational speech of young and elderly talkers. *J. Am. Acad. Audiol.* **7**, 212–218 (1996).
19. Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M. & Durlach, N. I. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J. Speech Lang. Hear. Res.* **39**, 494–509 (1996).
20. Picheny, M. A., Durlach, N. I. & Braida, L. D. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *J. Speech Lang. Hear. Res.* **29**, 434–446 (1986).
21. Ferguson, S. H. & Kewley-Port, D. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. (2007).
22. Krause, J. C. & Braida, L. D. Acoustic properties of naturally produced clear speech at normal speaking rates. *J. Acoust. Soc. Am.* **115**, 362–378 (2004).
23. Bottalico, P., Graetzer, S. & Hunter, E. J. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *J. Acoust. Soc. Am.* **139**, 2870–2879 (2016).
24. Koelewijn, T., Zekveld, A. A., Festen, J. M. & Kramer, S. E. Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* **33**, 291–300 (2012).
25. Zekveld, A. A., Kramer, S. E. & Festen, J. M. Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear.* **31**, 480–490 (2010).
26. Ishikawa, K., Li, H. & Coster, E. The effect of noise on initiation and maintenance of clear speech and associated mental demand. *J. Speech Lang. Hear. Res.* **66**, 4180–4190 (2023).
27. Sierra-Polanco, T., Cantor-Cutiva, L. C., Hunter, E. J. & Bottalico, P. Changes of voice production in artificial acoustic environments. *Front. Built Environ.* **7**, 666152 (2021).
28. ISO, I. 3382: Acoustics-Measurement of room acoustic parameters-Part 2: Reverberation time in ordinary rooms. 2008. *Geneva: ISO*
29. Pelegrín-García, D. & Brunskog, J. Speakers' comfort and voice level variation in classrooms: Laboratory research. *J. Acoust. Soc. Am.* **132**, 249–260 (2012).
30. Lezzoum, N., Gagnon, G. & Voix, J. Echo threshold between passive and electro-acoustic transmission paths in digital hearing protection devices. *Int. J. Ind. Ergon.* **53**, 372–379 (2016).
31. Lexicon. *MX400: Dual Stereo Reverb/Effects Processor. Datasheet,* https://lexiconpro.com/en-US/products/mx400#downloads.
32. Nilsson, M., Soli, S. D. & Sullivan, J. A. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* **95**, 1085–1099 (1994).
33. Kalikow, D. N., Stevens, K. N. & Elliott, L. L. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.* **61**, 1337–1351 (1977).
34. Hart, S. G. & Staveland, L. E. in *Advances in psychology* Vol. 52 139–183 (Elsevier, 1988).
35. Hershman, R., Milshtein, D. & Henik, A. in *Modern Pupillometry: Cognition, Neuroscience, and Practical Applications* 431–466 (Springer, 2024).
36. Hess, E. H. & Polt, J. M. Pupil size in relation to mental activity during simple problem-solving. *Science* **143**, 1190–1192 (1964).
37. Boersma, P. & Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. *Retrieved February* **3**, 2018 (2018).
38. De Jong, N. H. & Wempe, T. Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. Methods* **41**, 385–390 (2009).
39. Sweller, J. in *Psychology of learning and motivation* Vol. 55 37–76 (Elsevier, 2011).
40. Hodoshima, N., Arai, T. & Kurisu, K. in *Proc. International Congress on Acoustics.* 3632–3635 (Citeseer).
41. Bottalico, P., Łastowiecka, N., Glasner, J. D. & Redman, Y. G. Singing in different performance spaces: The effect of room acoustics on vibrato and pitch inaccuracy. *J. Acoust. Soc. Am.* **151**, 4131–4139 (2022).
42. Sirois, S. & Brisson, J. Pupillometry. *Wiley Interdiscip. Rev. Cognit. Sci.* **5**, 679–692 (2014).
43. Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E. & Smith, J. Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *J. Speech Language Hear. Res.* **61**, 1463–1486 (2018).

Author contributions

K.I. conceived and designed the study. K.I., P.B., S.M., H.L, and E.R. performed the experiments. H.L. and E.R. assisted with data organization. K.I. analyzed the data. K.I. wrote the manuscript and P.B., H.L., and E.R. revised it.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.