Emergence of Metacognitive Knowledge via Audible Pupil Size

Alwin de Rooij*
Tilburg School of Humanities and Digital Sciences (TSHD), Tilburg University, Tilburg, the Netherlands
alwinderooij@tilburguniversity.edu

Iris Wijers
TSHD, Tilburg University, Tilburg, the Netherlands
iris.wijers@outlook.com

Manon Marinussen
TSHD, Tilburg University, Tilburg, the Netherlands
manonmarinussen@outlook.com

ABSTRACT

Metacognitive knowledge is critical for adaptive behavior and depends on the ability to sense one's physiological signals. Some physiological signals, however, cannot be sensed yet carry critical information about one's thinking processes. The eye's pupils are an interesting example of this. Pupil size is typically inaccessible to the senses, yet it correlates with changes in attention and cognitive load. Using technology to map pupil size to audible sound in real-time, could therefore facilitate learning metacognitive knowledge. This was confirmed in a mixed between-within subject experiment, where participants (n=57) reported to be more able to use the sound to acquire metacognitive knowledge when using a real rather than a sham mapping from pupil size to sound during three thinking tasks. Specifically, participants using real audible pupil size were less likely to slightly disagree or be uncertain about their ability to use the sound to acquire metacognitive knowledge, but this difference was not found when participants reported high agreement. The contribution of this research is therefore that making pupil size audible can facilitate the emergence of new metacognitive knowledge.

CCS CONCEPTS

• **Human-centered computing** → Empirical studies in HCI.

KEYWORDS

Metacognitive knowledge, Pupil Size, Sensory Augmentation, Sonification

ACM Reference Format:
Alwin de Rooij, Iris Wijers, and Manon Marinussen. 2021. Emergence of Metacognitive Knowledge via Audible Pupil Size. In *European Conference on Cognitive Ergonomics 2021 (ECCE 2021), April 26–29, 2021, Siena, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3452853.3452870

1 INTRODUCTION

A Metacognition, the ability to think about one's own thinking processes, serves to adapt to changing environmental demands in order to maintain and improve task performance [7]. Effective metacognition critically depends on *metacognitive knowledge*, i.e., what one has learned about what thinking processes support or impede task performance in a given situation, and on control strategies, i.e., the execution of strategies to achieve and maintain the thinking processes thought to facilitate task performance [8]. Metacognitive ability predicts performance on learning, problem solving, and creative tasks [13], and can help compensate for potentially detrimental effects of low IQ on task performance [19].

Emerging evidence implies that *interoception*, the ability to sense physiological signals, supports the emergence of metacognitive knowledge [9]. Baroreceptors that sense the timing and strength of heartbeats, e.g., can drive the changes in the noradrenergic activity in the brain that makes increased attentional focus toward stimuli in the direct environment possible [20]. Such relationships between physiology and thinking can be sensed, learned as part of metacognitive knowledge, and acted upon in future situations 17].

Some physiological signals, however, tend to fall outside one's interoceptive awareness. Yet these signals may carry critical information about one's thinking processes. A key example are *the eye's pupils.* Pupil size correlates with changes in amongst others attentional focus and cognitive load [3]. Information that could at times be critical for effective metacognition but is not accessible as such [15]. This, combined with advances in consumer grade sensing technology, invites speculation about how technology can leverage physiological signals such as pupil size to support the emergence of metacognitive knowledge.

In the present paper, it is conjectured that *audible pupil size*, i.e., making changes in pupil size available to the auditory sense in real-time, could be such a novel way to support the emergence of metacognitive knowledge. In what follows, this conjecture is unpacked in more detail based on a predictive processing account of learning and sensory augmentation. The method and results of a mixed between-within subject experiment are presented that test whether real audible pupil size, compared sham audible pupil size, can indeed support the emergence of metacognitive knowledge. Thereafter the study's results are discussed, and two design challenges are prosed as future steps that need to be taken to make practical application possible.

1.1 Metacognitive knowledge

Metacognitive knowledge, one's learnings about what thinking processes support or impede performance in a given situation, is critical for the ability to think about one's own thinking processes – and for acting upon these thoughts to maintain or even improve one's task performance [7]. Taking a predictive processing perspective on learning, one could argue that metacognitive knowledge serves

*Corresponding author.

ECCE 2021, April 26–29, 2021, Siena, Italy

ACM ISBN 978-1-4503-8757-6/21/04…$15.00
https://doi.org/10.1145/3452853.3452870

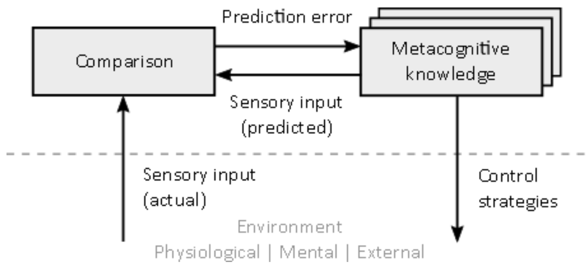

Figure 1: A model of how metacognitive knowledge is learned.

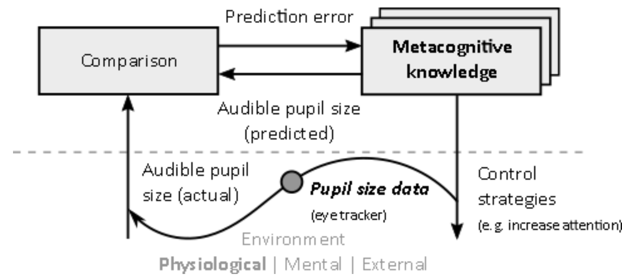

Figure 2: A model of how audible pupil size may support the emergence of metacognitive knowledge.

to generate predictions about the (sensory) input that metacognitive control strategies should elicit to facilitate task performance [8] (Figure 1). Comparison between the predicted and the actual (sensory) input that results from an ongoing thinking process enables learning and updating of metacognitive knowledge. That is, when a discrepancy between the predicted and actual (sensory) input occurs (prediction error), this discrepancy is encoded in memory, updating or generating new metacognitive knowledge [18]. Sustained focused attention, for example, may coincide with a prediction error that signals better task performance in the analytical, but not the creative aspects of a task, encoding knowledge about differences in the efficacy of sustained attention for analytical and creative thinking [7].

The *contents of metacognitive knowledge* emerge from the contingencies between ongoing thinking processes and (sensory) input present at the moment that a prediction error occurs [18]. These contents can be *physiological* (e.g., variability in heart rate) [2], mental (e.g., feelings of effort) [7], and external (e.g., visual task features) [8]. Prediction error at any of these levels can drive updating or the generation of new knowledge about what thinking processes support or impede task performance, and under what circumstances. For example, while visual cues indicate the type of task one works on, actually sensed heart rate variability, which correlates with effort feelings, may be higher than predicted, thereby updating metacognitive knowledge about how much felt effort is needed during this type of task.

Interoception, the ability to sense physiological signals that originate from within the body, therefore also plays a key role in learning metacognitive knowledge [2]. By making it possible to sense physiological signals, correlations between these signals and thinking can be learned and actuated. For example, baroreceptors that sense the timing and strength of heartbeats drive changes in noradrenergic activity in the central nervous system, which focuses attention [20]. In turn, noradrenergic activity may reciprocally regulate the increases in heart rate variability needed to maintain that same attentional focus [20]. Interoception makes it possible to evaluate discrepancies between the physiological signals expected during a thinking process, and the actual physiological signals that emerge during that process. Discrepancies that drive learning of metacognitive knowledge (as described above).

However, some physiological signals tend to fall outside one's interoceptive awareness [15], yet may also carry critical information about one's thinking processes [3]. This invites speculation about whether (new) metacognitive knowledge can be brought about by using a *technology* that makes otherwise inaccessible physiological signals that carry information about one's thought processes, accessible.

1.2 Audible pupil size

The *eye's pupils,* dark circular openings in the irises that dilate and constrict via (de)activation of radial and circular muscles surrounding each pupil [3], are a good example of a typically inaccessible physiological signal that carries information about one's thinking processes [15]. The pupils regulate the amount of light that falls on the retina. However, pupil size also rapidly tracks changes in thinking related processes. Under isoluminant conditions, correlations between activity in the locus coeruleus, a noradrenaline rich area located in the pons with projections into amongst others the prefrontal cortex, can be observed [3]. Noradrenaline modulates the cortical signal-to-noise ratio, which implements the allocation of cognitive resources to task-relevant and task-irrelevant stimuli. Fast changing task-evoked pupil size positively correlates with maintaining focus to task-relevant stimuli, and relatedly mental effort and analytical thinking [16]. Whereas slower changing baseline pupil size positively correlates with the alertness to task-irrelevant stimuli, and relatedly distractibility, flexibility, and creative thinking [16]. Correlations of pupil size with task difficulty, depth of thinking, stress, and cognitive load can be explained in a similar manner [3].

Building upon the sensory augmentation paradigm [17], it can be conjectured that a real-time mapping between pupil size, measured with an eye-tracker, to audible sound, can bring about new metacognitive knowledge (Figure 2) [15]. Whenever contingencies between ongoing thinking processes and (sensory) input happen when a prediction error occurs [18], contingencies between such *audible pupil size* and one's thinking processes can be learned just like any other input from the physiological, mental, or external environment. For example, if we assume that pupil size positively correlates with attentional focus during a mental arithmetic task [3], metacognitive knowledge may be learned via the support of a mapping of pupil size to the volume of a sound in real-time [15]. That is, when a user hears these sounds while engaging in a task that requires focus, contingencies between lowered sound volume, lowered attentional focus, and at the moment the task is going less well than expected (prediction error) can be learned as new metacognitive knowledge.

Recent work from the human-computer interaction community provides some insight into this kind of approach. Correlational studies by Ehlers and colleagues showed that by making pupil size visible, thoughts that elicit changes in arousal could be used by 50% of their participants to gain some control their visible pupil size [4–6]. Complementarily, de Rooij and colleagues showed in another correlational study that by making pupil size audible, via a mapping of pupil size to the volume of a sine wave, participation in a mathematics task elicited associations between the audible pupil size and the depth of thinking, the moment of focusing, and difficulty for 55% of the participants [15]. This increased 75% after instructing participants about the existing relationship between pupil size and their thinking. This suggests that users of audible or visible pupil size are able to observe contingencies between the pupil size signals and their thinking processes, a condition necessary for the emergence of metacognitive knowledge via audible pupil size.

1.3 Research question

These conjectures and related previous work suggest that audible pupil size essentially enables users to hear their own noradrenaline system at work while engaged in a thinking process. Speculatively, this may enable users to integrate the wide spectrum of correlations of pupil size with different thinking-related processes into the contents of new metacognitive knowledge. Given the novelty of these conjectures and the scarcity of mostly remotely related work, however, these conjectures present an open scientific and design problem. As a first step toward addressing this open problem, the following research question will be answered: *Does audible pupil size facilitate the emergence of metacognitive knowledge?*

2 METHOD

To answer the research question an experiment with a mixed between-within-subject design was conducted.

2.1 Participants

Fifty-eight people participated in this study. Data from one participant was not used due to a failure to follow the procedure. Data from the remaining fifty-seven participants (M_{age} = 24.10, SD_{age} = 3.23, $range_{age}$ = [19-34], 13 Male, 44 female) was used in the analysis. Participants were recruited via the participant recruitment system of the communication and information sciences program of Tilburg University. The participants received course credit in exchange for their participation.

2.2 Materials and measurements

The between-subject conditions were the randomly assigned exposure to real or sham audible pupil size during three thinking tasks (Figure 3). This design was chosen because 1) related work has produced evidence only using correlational designs [4–6, 15] and 2) other research has shown that making physiological signals perceivable for self-regulatory ends (biofeedback) often does not outperform placebo [9]. People are masters at pattern finding, and thus one would expect that sham audible pupil size could also lead to perceiving (illusory) contingencies between pupil size and the participants thinking [15]. *Real audible pupil size* was a real-time mapping from pupil size to the volume of an audible sound. An

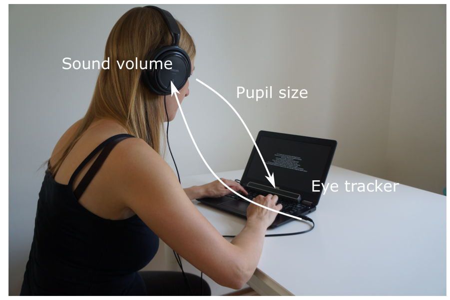

Figure 3: Experimental setup.

EyeTribe eye-tracker was used to capture pupil size. The pupil size data was streamed to the Cycling '74 MAX7 environment via the OSC protocol, where it was mapped to volume changes of a continuously playing sine wave (440 Hz). Prior to use the minimum and maximum pupil size of the user was recorded to normalize the signal. Signal loss, e.g. due to blinking, was handled by keeping the last known value until a new value was passed. The resulting signal was then mapped directly to volume. *Sham audible pupil size* was a generated random variation in loudness that sounded similar to the real audible pupil size. A sine wave with a frequency range of [0.25 Hz, 1.00 Hz] and a sine wave with a frequency range of [1.00 Hz, 2.00 Hz] were multiplied and normalized between 0 and 1. The frequency ranges changed at random times between [1000 ms, 7500 ms]. The resulting signal was used to modulate the volume of the audible 440 Hz sine wave. The researcher subjectively decided what settings were similar to the real audible pupil size, resulting in sonically similar volume changes in the same sound. The protocol needed for setting up the eye-tracker in the real audible pupil size condition was also applied in the sham condition.

The within-subject conditions were *three thinking tasks* that are known to elicit changes in pupil size. Prior to the tasks, participants were told that they would be hearing sounds during the thinking tasks, that these were generated based on their pupillary activity, and therefore could correlate with their thinking processes. Furthermore, they were instructed to try to use the sounds to learn something about their thinking during the tasks. In task 1, a reading task, participants read twelve statements that varied from syntactically simple sentences with subject-relative constructions ("The man that attacked the burglar protested the arrest") and syntactically complex sentences with object-relative constructions ("The man that the burglar attacked protested the arrest.") [14]. In task 2, a verbal problem-solving task, participants solved twelve problems that varied randomly from easy to difficult problems [1]. In task 3, a mathematics task, participants solved twenty multiplication questions mentally [12] that varied in difficulty (e.g., "5 x 5" versus "13 x 17"). Note that these tasks served to elicit variation in the contingencies between pupil size and thought, which is needed to learn metacognitive knowledge. Thus, subsequent analysis is focused on their combined effect rather than on testing specific differences between or within the thinking tasks.

To assess the effects of real vs. sham audible pupil size on the emergence of metacognitive knowledge, participants rated the following Likert scale: "*I was able to use the sound to acquire knowledge* *about myself and my cognition"* (variable: metacognitive knowledge learned). To ensure that the metacognitive knowledge learned could be attributed to the contingencies that the participants perceived between the audible pupil size and their thinking processes they also rated the following scale: *"I observed a correlation between my own cognitions or cognitive behaviour and changes in the sound"* (variable: relationship sound-cognition perceived). Finally, to check whether differences related to metacognitive knowledge could not be inadvertently attributed to differences in the ability to perceive changes in the sound at all, the participants also rated: *"I heard the sound changing during this task"* (variable: sound perceived). All scales were 7-point Likert scales (1 = Not agree at all, 7 = Completely agree) and were rated after each of the three tasks. Other data were collected as well to inform decisions on future work but are not published in the present paper.

2.3 Procedure

Upon arrival participants sat down behind a laptop in an isoluminant environment. Dimmed lights were used with a low brightness screen to reduce the pupil's light reflex. Participants were introduced to the study, signed informed consent, and filled in a questionnaire about their socio-demographics. The researcher provided further instructions and explained the relationship between the sounds the participants would be hearing, pupil size, and their thinking. The eye-tracker was put on and calibrated (9-point calibration). The minimum and maximum pupil size was captured by darkening and brightening the laptop screen. Participants then put on headphones and adjusted the overall sound volume to a comfortable but clearly audible level. The three thinking tasks were then executed. Prior to each thinking task, specific instructions on how to execute that task were presented. After each task participants filled in the questionnaire used to assess the emergence of metacognitive knowledge. The study took up to 45 minutes. The researcher was blind to the experimental conditions.

2.4 Data analysis

The dataset was analysed using R statistics version 4.0.2. To provide insight into the general characteristics of the dataset histograms were generated (Figure 3) and the descriptive statistics (Table 1, left) and sample size weighted between subject Spearman correlations (Table 1, right) were calculated. Spearman correlations were chosen over Pearson correlations due to the deviation from normality displayed in the histograms (Figure 4). Sample size weighting was applied to account for repeated measures. Quantile linear mixed modelling (QLMM) was used to help answer the research question [10, 11]. The independent variable was sham (coded: 1) vs. real (coded: 2) audible pupil size. Data were grouped by participant to account for repeated measures. Three such models were calculated with metacognitive knowledge learned (model 1), relationship sound-cognition perceived (model 2), and sound perceived (model 3) as the dependent variable. The choice for using LQMM was based on the following diagnostic tests. First, visual inspection of the histograms showed that all variables' distributions departed from normality (Figure 3). Second, tests using generalised linear mixed models (GLMM) with nonparametric gamma and beta (with scaled data) distributions yielded significant Kolmogorov-Smirnov tests. Thus, GLMMs assumptions were not met. Third, quantile regressions between the predicted data GLMMs and the observed data at the quantiles τ = .25, τ = .50, and τ = .75 quantile, showed a significant deviation from linearity. Note that in quantile regression and relatedly QLMM, τ are the quantiles used to calculate the regression and QLMM models' intercepts. These diagnostic tests suggested multimodality in the variables metacognitive knowledge learned and relationship sound-cognition perceived. Multiple unmeasured causes may therefore be responsible for the variation observed in these variables. QLMM is a suitable method in such cases because it enables exploration of where experimental manipulations may have effects on a multimodal variable, and where not, while accounting for repeated measures [11]. Point estimates at quantiles τ = [.10, .90, step size = .10] were calculated, which is a standard resolution for exploratory testing [10]. Twenty Monte Carlo simulations were ran to ensure high accuracy [10]. The maximum number of iterations was fifty, as initial tests suggested no substantial improvements in model convergence well before the fiftieth iteration finished.

3 RESULTS

To provide insight into the characteristics of the dataset histograms were generated (Figure 4) and descriptive statistics and the sample size weighted between subject Spearman correlations were calculated (Table 1).

To results of model 1 showed a significant difference between sham vs. real audible pupil size for metacognitive knowledge learned (Figure 5, left; Table 2, top). This difference was found at the .21 intercept (τ = .10), b = .95, p = .010, 95% CI[.23 1.68], the 3.06 intercept (τ = .50), b = .92, p = .025, 95% CI[.12 1.73], the 3.14 intercept (τ = .60), b = .86, p = .041, 95% CI[.04 1.68], and the 4.01 intercept (τ = .80), b = .93, p = .028, 95% CI[.10 1.76]. These findings suggest that participants exposed to real audible pupil size were less likely to report not being able to acquire metacognitive knowledge at all than participants exposed to sham audible pupil size (compare intercept at τ = .10 with Likert score 1 = Not agree at all). Moreover, participants exposed to real audible pupil size were less likely to slightly disagree or be unsure about being able to use the sound to acquire metacognitive knowledge (compare intercepts at τ = .50, .60, .80 with Likert scores 3 = slightly disagree, and 4 = neither agree nor disagree).

The results of model 2 also showed a significant difference between sham vs. real audible pupil size for the relationship sound-cognition perceived (Figure 5, middle; Table 2, middle). This effect was found at the 3.08 intercept (τ = .40), b = .91, p = .027, 95% CI[.10 1.71], and 4.17 intercept (τ = .70), b = .82, p = .019, 95% CI[.14 1.50]. This suggests that participants exposed to real audible pupil size were less likely slightly disagree or be uncertain about perceiving a relationship between the sounds and their thinking (compare intercepts at τ = .40, .70 with Likert scores 3 = slightly disagree, and 4 = neither agree nor disagree).

The results of model 3 showed one significant difference between sham vs. real audible pupil size for sound perceived (Figure 5, right; Table 2, bottom). This was at the 6.74 intercept (τ = .50), b = -.84, p = .022, 95% CI[-1.55 -.12]. Note that the intercept of this one significant difference tends to coincide with values well above the intercepts (Table 2) where significant differences for metacognitive

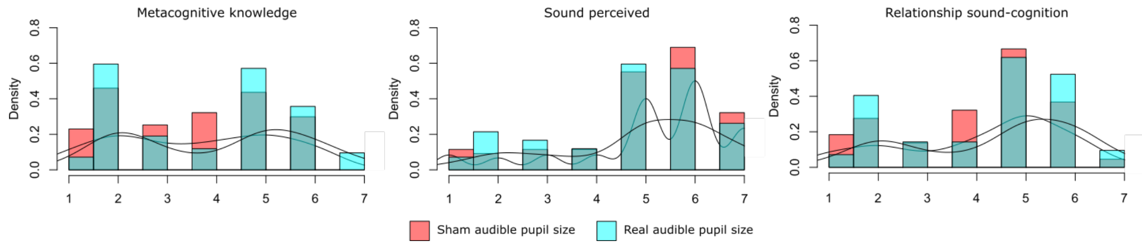

Figure 4: Histograms including density lines split by condition (red= sham audible pupil size, blue = real audible pupil size). X-axis represents the number of ratings for each point on the Likert scales. Y-axis represents density.

Table 1: Descriptive statistics and sample size weighted between group correlations.

		Mean	SD	Spearman correlations		
				1.	2.	3.
1. Metacognitive knowledge	R	3.99	1.75	-		
	S	3.59	1.66			
2. Sound perceived	R	4.84	1.66	.73***	-	
	S	5.13	1.62			
3. Relationship sound-cognition	R	4.39	1.69	.77***	.76***	-
	S	4.15	1.65			

Note. Correlation data (right) are Spearman correlation coefficients and 95% confidence intervals (between square brackets). SD = Standard deviation, IQR = Inter Quartile Range. R = Real audible pupil size, S = Sham audible pupil size. *** $p < .001$.

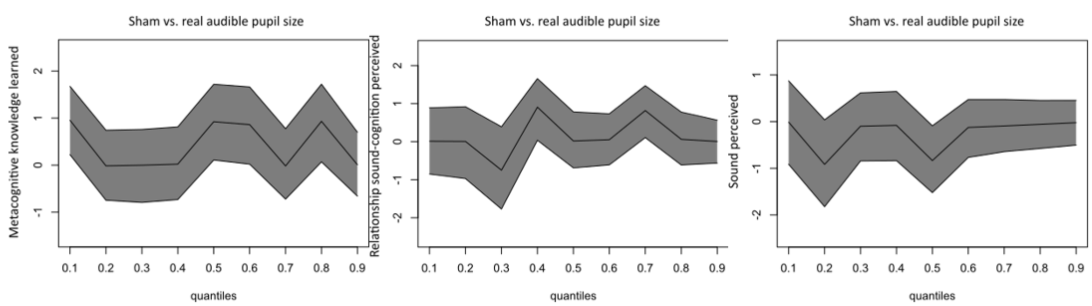

Figure 5: Point estimates (middle line) and 95% confidence interval (grey area) of real relative to sham audible pupil size for the variables metacognitive knowledge, relationship sound-cognition perceived, and sound perceived (y-axis) for QLMM results at quantiles $\tau = [.10\ .90]$ (x-axis).

knowledge learned and relationship sound-cognition perceived were found. Differences, if any, in the ability to perceive changes in the sound therefore unlikely explain the effects of audible pupil size on metacognitive knowledge learned or on the perception of a relationship between the sounds and the participants' thinking.

4 DISCUSSION

In this paper, it was tested whether audible pupil size facilitated the emergence of metacognitive knowledge.

4.1 Emergence of metacognitive knowledge via audible pupil size

The results showed that real audible pupil size, when compared to sham audible pupil size, could lead participants to report being more able to use the sound to acquire metacognitive knowledge. The findings suggested that participants exposed to real audible pupil size were less likely to report not being able to use the sound to acquire any metacognitive knowledge at all than participants exposed to sham audible pupil size. Moreover, participants exposed to real audible pupil size were less likely to slightly disagree or be uncertain about their ability to use the sound to acquire metacognitive knowledge. Importantly though, no significant differences were found between real and sham audible pupil size when participants reported with more confidence in their ability to use the sound to acquire metacognitive knowledge. When there were differences, this was likely driven by the ability to perceive contingencies between the sound and the participants' thinking – a theoretical condition for metacognitive knowledge to emerge. This because 1) at the same intercepts, where differences between real and sham

Table 2: Effects of real versus sham audible pupil size on metacognitive knowledge learned (top), relationship sound-cognition perceived (middle), and sound perceived (bottom).

	Metacognitive knowledge learned								
	τ = .10	τ = .20	τ = .30	τ = .40	τ = .50	τ = .60	τ = .70	τ = .80	τ = .90
Intercept	0.21 (.61)	2.13*** (.61)	2.15** (.67)	3.04*** (.62)	3.058*** (.62)	3.14**** (.67)	4.97*** (.68)	4.01*** (.67)	5.89*** (.55)
Condition	.95*(.37)	-.02 (.38)	-.01 (.40)	.02 (.39)	.92* (.41)	.86* (.42)	-.02 (.38)	.93* (.10)	.01 (.35)
Sigma	.12	.23	.31	.35	.37	.36	.33	.24	.14
AIC	631.71	628.96	637.06	619.73	628.61	624.77	634.19	644.06	651.00
	Relationship sound-cognition perceived								
Intercept	2.07** (.69)	2.15** (.79)	4.74*** (.91)	3.08*** (.62)	4.88*** (.68)	4.88*** (.61)	4.17*** (.59)	5.83*** (.55)	5.92*** (.47)
Condition	.01 (.44)	.01 (.48)	-.75 (.56)	.91* (.41)	.02 (.38)	.05 (.34)	.82* (.35)	.06 (.36)	.01 (.29)
Sigma	.16	.30	.39	.44	.45	.43	.37	.27	.15
AIC	69.84	692.80	672.68	663.08	655.92	643.65	645.06	644.73	635.36
	Sound perceived								
Intercept	2.18** (.73)	5.01*** (.82)	5.13*** (.64)	5.13*** (.62)	6.74*** (.59)	6.08*** (.49)	6.08*** (.45)	6.06*** (.43)	6.94*** (.40)
Condition	-.01(.46)	-.92 (.48)	-.10 (.37)	-.10 (.38)	-.84* (.37)	-.12 (.32)	-.09 (.29)	-.06 (.26)	-.02 (.24)
Sigma	.15	.27	.36	.40	.42	.41	.35	.25	.15
AIC	698.18	667.26	656.38	637.27	627.10	619.09	602.60	593.25	610.98

Note. Data are point estimates and standard errors (between parentheses) at τ = [.10, .90, step size = .10]. τ is the quantiles used as an intercept. Condition is the effects of real compared to sham pupil size. AIC = Akaike Information Criterion. * p < .050, ** p < .010, *** p < .001.

audible pupil size were found for the ability of use the sound to acquire metacognitive knowledge, differences were also found for the ability to perceive a relationship between the sound and the participants' thinking, and 2) there was a positive correlation between these variables. Here also, however, no significant differences were found between real and sham audible pupil size when participants reported with more confidence that they perceived a relationship between audible pupil size and their thinking. Finally, the differences that were found could not be explained by more general differences in the ability to perceive changes in the sound between the real and the sham audible pupil size. One significant difference was found, but data at the intercepts where significant differences were observed the ability to use the sound to learn metacognitive knowledge and the perceived relationship between sound and thinking, did not align with the data at the intercepts at which the ability to perceive the sound differed between the real and the sham audible pupil size. This finding supports the validity of the presented results.

4.2 Limitations

There are, of course, also several limitations that need to be taken into account when interpreting and building upon the presented results. Two of these are presented in more detail below. First, some of the methodological choices threaten the result's construct validity. Central to this was the use of self-report. Although self-report simplifies capturing whether participants had been able to use the sounds to acquire a possible broad range of metacognitive knowledge, which we were reluctant to narrow down beforehand, this has the obvious drawback of lacking an objective measurement of the metacognitive knowledge acquired. Similarly, whether actual contingencies between pupil size and thinking processes occurred was not checked objectively based on pupil size recordings. Using single items, rather than multiple also introduces uncertainty about whether all participants interpreted these items correctly. Especially because their understanding of the items was not verified, nor were the questions used previously validated in any way. In addition, the items used demand quite some insight from the participants both linguistically and conceptually. Second, the multimodality of the data complicated the interpretation of the causal relationships found between real vs. sham audible pupil size and the emergence of metacognitive knowledge and related measured variables. This threatens the study's internal validity. Interpretation, though informed by previous work, therefore remains speculation (see section 4.3). Thirdly, it could be tempting to assume that the found emergence of metacognitive knowledge could translate into improved future task performance, c.f. [7, 8]. The results of the present study, however, cannot be used to substantiate such claims – nor was it intended as such. Therefore, no conclusions can be drawn about, say, whether audible pupil size could be used to achieve performance benefits on thinking tasks. It may well be possible that hearing continuous sound during a thinking task is detrimental to such an extent that it mitigates any performance benefits that may occur from any metacognitive knowledge learned. Rather, the present study should be seen within the context of a "what if?" question. Paving the way for further research.

4.3 Design challenges: Steps toward practical application

The findings also point to at least two major design challenges that need to be addressed when working toward practical applications that involve audible pupil size. First, a substantial subset of users may not be able to make use of audible pupil size. The findings in the present study echo previous work [15]. There, only 55%-75% of the participants were able to perceive contingencies between audible pupil and their thinking process. This may explain a peak in the data for participants that reported to not be able the use the sound to acquire metacognitive knowledge. It must be noted that participants that used real audible pupil size were less likely to report not being able to acquire metacognitive knowledge at all, than participants exposed to sham audible pupil size. Suggesting that some of the inability of participants to make effective use of the sound, could perhaps be mitigated by further design driven research. One design challenge would therefore be to maximise the percentage of users that is able to perceive contingencies between pupil size and their thinking. This could partially be addressed by improving the mapping from pupil size to sound, exploring the efficacy of other modalities than sound, and investigating its effects in a wider range of thinking tasks than was done in the present study. Second, there may be uncertainty about the degree to which real audible pupil size tends to cause illusory contingencies between the sounds and the participants thinking processes, possibly causing the emergence of erroneous metacognitive knowledge. Surely, contingencies between the sounds and thinking may happen at random, causing some participants to pick up on these. This may have caused some that were exposed to the sham audible pupil size to report with confidence that they were able to use the sound to acquire metacognitive knowledge. Leading to no significant difference between sham and real audible pupil size when more confidence was reported about having learned metacognitive knowledge and having perceived a relationship between audible pupil size and their thinking process. However, this also introduces uncertainty about the degree to which participants in the real audible pupil size perceive such illusory contingencies. This casts doubt over previous correlational studies as well [3–5, 15]. Therefore, the second major design challenge is preventing the emergence of illusory contingencies between the sounds and a user's thinking, therewith preventing the subsequent emergence erroneous metacognitive knowledge.

4.4 Contribution statement

Given these findings, it can be concluded that the presented study contributes empirical evidence for the conjecture that audible pupil size can facilitate the emergence of new metacognitive knowledge. However, future practical application will depend on solving several key design challenges.

REFERENCES

[1] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological bulletin, 91, 276.
[2] Hugo D. Critchley and Sarah N. Garfinkel. 2018. The influence of physiological signals on cognition. Current Opinion in Behavioral Sciences, 19, 13-18.
[3] Maria K. Eckstein, *et al.* 2017. Beyond eye gaze: What else can eye tracking reveal about cognition and cognitive development?. Developmental cognitive neuroscience, 25, 69-91.
[4] Jan Ehlers, Juliane Georgi, and Anke Huckauf. 2014. Improving voluntary pupil size changes for HCI. In Proceedings of the 8th international conference on pervasive computing technologies for healthcare (pp. 343-346). ICST.
[5] Jan Ehlers, Christoph Strauch, Juliane Georgi, and Anke Huckauf. 2016. Pupil size changes as an active information channel for biofeedback applications. Applied psychophysiology and biofeedback, 41(3), 331-339.
[6] Jan Ehlers, Christoph Strauch, and Anke Huckauf. 2018. Training facilitates cognitive control on pupil dilation. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (pp. 67-72). ACM.
[7] John H. Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American psychologist, 34, 906.
[8] Stephen M. Fleming and Raymond J. Dolan. 2012. The neural basis of metacognitive ability. Phil. Trans. R. Soc. B, 367, 1338-1349.
[9] Jerome Frank. 1982. Biofeedback and the placebo effect. Biofeedback and self-regulation, 7, 449-460.
[10] Christian E. Galarza. 2020. Package 'qrLMM'. Retrieved on June 2nd from https://cran.r-project.org/web/packages/qrLMM/qrLMM.pdf
[11] Christian E. Galarza, Victor H. Lachos, and Dipankar, Bandyopadhyay. 2017. Quantile regression in linear mixed models. Statistics and its Interface, 10, 471.
[12] Eckhard H. Hess and James M. Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. Science, 143, 1190-1192.
[13] Norbert Jaušovec. 1994. Metacognition in creative problem solving. Ablex Publishing.
[14] Marcel A. Just and Patricia A. Carpenter. 1993. The intensity dimension of thought: pupillometric indices of sentence processing. Canadian Journal of Experimental Psychology, 47, 310.
[15] Alwin de Rooij, Hanna Schraffenberger, and Mathijs Bontje. 2018. Augmented metacognition: Exploring pupil dilation sonification to elicit metacognitive awareness. In Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction (pp. 237-244). ACM.
[16] Alwin de Rooij and Ruben D. Vromans. 2018. Noradrenergic modulation of creativity: evidence from pupillometry. Creativity Research Journal, 30, 339-351.
[17] Frank Schumann and J. Kevin O'Regan. 2017. Sensory augmentation: Integration of an auditory compass signal into human perception of space. Scientific reports, 7, 42197.
[18] Anil K. Seth. 2013. Interoceptive inference, emotion, and the embodied self. Trends in cognitive sciences, 17, 565-573.
[19] Lee H. Swanson. 1992. The relationship between metacognition and problem solving in gifted children. Roeper Review, 15, 43-48.
[20] Julian F. Thayer, *et al.* 2012. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neuroscience & Biobehavioral Reviews, 36, 747-756.