Assessment of a modified temporal bisection task with responses on a continuous visual scale

Ivan Quan[1]

(Rebekka Lagacé-Cusiac[2])

(Dr. Jessica Grahn[2])

[1]Department of Physiology & Pharmacology, [2]Department of Psychology, [1]Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

Table of Contents

1 Abstract

Temporal perception is a fundamental part of our experiences, and is intrinsically connected to our sensations, actions, and cognitions. It can be studied with the temporal bisection task, which involves training participants on a short and long reference tone before presenting intermediate tone durations and tasking participants to classify them as “short” or “long”. However, the temporal bisection task is limited in that it can only capture a binary response, and is not able to accordingly determine the human capacity to quantify relative durations. Previous studies have also not determined the reliability of this task alongside its use. We performed a modification to classic bisection, in which participants responded to tones by placing the tone along a continuous slider between references. We hypothesized that humans could quantify relative durations in the modified bisection task, and that this task would reach adequate reliability for use in future research.

Participants (n = 28) completed the classic and modified bisection tasks twice across two sessions. Results from the modified task showed that true accuracy scores were significantly better than scores from randomized responses. Results also indicated excellent internal-consistency and test-retest reliability for the modified task, while the classic task showed poor reliability in both. Correlations between accuracy scores from the classic and modified tasks demonstrated fair convergent validity between scores.

These results displayed that humans are able to quantify relative durations in the temporal bisection task. The assessed reliability of this task also indicated a level of robustness suitable for use in future studies. The modified temporal bisection task could be used for extracting accuracy scores to assess individual or group differences in temporal perception, or to study human quantification of relative durations. In the future, replicating this preliminary study with a larger sample size could provide further evidence to support its use.

Keywords

Relative time perception, temporal discrimination, ratio processing, continuous measures, reliability

2 Introduction

2.1 Temporal Perception

Temporal perception is an essential part of nearly everything we experience and is innately linked with the perception of sensory information (Wittmann, 2009). The brain continuously processes temporal information by measuring durations, recalling temporal information stored in memory, and comparing time durations to inform both our conscious and subconscious operations. From organizing daily schedules, to producing motor outputs, and even the comprehension and generation of speech, cognitions underlying time perception are critical for many key processes (Lindbergh and Kieffaber, 2013). A difficulty with studying temporal cognition is that its sensations cannot be traced from an obvious source organ (Matthews and Meck, 2014). However, many studies surrounding temporal perception are able to discover how we process temporal information using psychophysical approaches (García-Pérez, 2014). This involves analyzing sensations by making quantitative measurements about perceptions while varying stimulus properties (Read, 2015). Psychophysics is concerned with examining how sensory experiences vary based on changing stimuli, to ultimately make inferences about the cognitive processes behind sensation and perception (Read, 2015). A psychophysical approach towards studying time perception allows researchers to relate physical stimuli to mental perceptions and uncover processes behind temporal perception without directly involving neurocircuitry.

2.2 Temporal Bisection Task

The temporal bisection task is a well-established psychophysical method used to study temporal discrimination and the storage of time information in memory (Lindbergh and Kieffaber, 2013). Conventionally, this task first involves training participants on two reference stimuli labelled “short” and “long” (Allan and Gibbon, 1991). Tones of constant frequency are typically used as stimuli, and these are presented to participants to hold in memory. Following reference training, participants are then presented with tones of intermediate duration (Wearden and Ferrara, 1995). Participants are tasked with the classification of each presented tone in a binary two-alternative choice as either the “short” or “long” reference duration based on their perception of the presented stimulus duration (Kopec and Brody, 2010).

2.3 Bisection Point

A key measure arising from the temporal bisection task is the bisection point (Wearden, 1991). When plotting the probability of responding “long” against the presented stimulus duration, participants are more likely to respond “long” as stimulus length increases. The bisection point refers to the stimulus duration when this probability reaches 50%, where participants are equally as likely to respond “short” or “long” (Allan and Gibbon, 1991). Many studies of the temporal bisection task have used different parameters for reference duration magnitude and the ratio between reference durations, and their effects on the bisection point have been extensively analyzed (Kopec and Brody, 2010). Generally, the bisection point lies slightly below the arithmetic mean of the two reference durations (Kopec and Brody, 2010). Although the magnitude of reference durations does not have a large effect on the location of the bisection point, when the ratio between the reference durations increases, the bisection point tends to fall increasingly below the arithmetic mean of the reference durations (Kopec and Brody, 2010).

2.4 Measuring Temporal Perception Ability

While the temporal bisection task has primarily been used to study internal processes of time perception, scores derived from this task have also been used to measure group differences in temporal perception between clinical populations (Liu et al., 2022). When plotting proportions of “long” responses against stimulus duration, a measure of the steepness of the slope encompassing the central region including the bisection point has been used as an indication for temporal discrimination ability (Elvevåg et al., 2003). A steeper slope around the bisection point indicates that a participant would notice a smaller change in stimulus duration to determine more precisely which presented stimuli were closer to their respective references. This score has been linked to conditions such as cerebellar degeneration and schizophrenia, finding that people affected by these disorders had impaired judgements in timing (Nichelli et al., 1996; Elvevåg et al., 2003).

2.5 Modeling Responses to the Temporal Bisection Task

Participant responses to the temporal bisection task have been modeled as a two-step process in decision-making (Kopec and Brody, 2010). Internal representations of reference durations can be modeled as normal distributions centered around the perceived length of learned durations, with standard deviations proportional to the magnitude of these learned durations (Church and Gibbon, 1982). The likelihood of recognizing a stimulus of certain duration as the learned reference is represented by the height of each distribution (Church and Gibbon, 1982). With two normal distributions representing “short” and “long” reference durations, there may exist a span in between the distributions where a stimulus is not recognized as either reference. Accordingly, a two-step model for decision-making arises. First, a participant determines if a presented stimulus is mentally represented as either reference duration (Kopec and Brody, 2010). Upon determining that a presented stimulus is a reference duration, they would give an answer at this point. However, if the stimulus was perceived to be neither of the reference durations, they would proceed to the second step to compare the relative distance between the presented stimulus and reference durations (Kopec and Brody, 2010). Participants would then answer based on which reference appeared closer. Because of the inherently larger standard deviation of the long reference compared to the short reference in this model, a larger proportion of presented stimuli durations will be immediately recognized as the long reference (Church and Gibbon, 1982). As a result, stimuli represented as intermediate durations will tend towards more short responses due to the gambler's fallacy, driving the belief that a previous "long" response decreases the probability of another "long" response (Kopec and Brody, 2010).

2.6 Limitations of the Temporal Bisection Task

While this two-step model is a useful interpretation of how our decision-making process in the temporal bisection task can produce effects seen in task data, it provides only a limited illustration of how we perceive intermediate durations (Lindbergh and Kieffaber, 2013). Although mental processes behind recognizing and comparing intermediate durations have been elucidated when modeling participant responses to the temporal bisection, exactly how these intermediate durations are internally represented has not been measured in a precise manner. Because participants are forced to respond in a two-alternative "short" or "long" decision, previous analyses have not been able to accordingly deduce the capability of quantifying relative durations in the temporal bisection task. In timing estimation tasks, instead of using a binary measure, temporal judgements can also be recorded on a continuous scale. For ratio judgments on empty time intervals with auditory stimuli of interest surrounded by two warning signals, estimates of relative timing have been made by placing marks on a bounded linear scale to indicate the relative location of stimuli (Nakajima, 1987). For empty-interval auditory stimuli consisting of a single tone, or a pattern of three tones, people have been able to make accurate judgements of relative timing by placing markings on a bounded line (Nakajima, 1987; Lagacé-Cusiac et al., 2023).

2.7 Current Study

The current study investigated a modification to the temporal bisection task. As an alternative to a binary decision between “short” or “long”, participants responded to presented stimuli instead by placing the intermediate duration along a visual continuous sliding scale bounded by the two reference durations. This study also analyzed the internal-consistency and test-retest reliability of both the classic and modified bisection tasks. Although studying the reliability of psychophysical measures is essential towards making robust inferences to mental processes, analyses of reliability have not generally been conducted alongside its use, which is an issue pervasive to many psychological measures (Parsons et al., 2019). Because a measurement with subpar reliability is inconsistent in obtaining similar results under similar conditions, measurement error can play a large role in psychophysical task outcomes (Parsons et al., 2019). Determining the reliability of temporal bisection measures can give insight on the robustness of the inferences made to temporal perception and can inform the use of the modified bisection task in the future.

The current study was conducted with two primary objectives. We aimed to determine if humans can quantify relative durations in a modified temporal bisection task with responses on a visual analog scale. Additionally, we aimed to determine the reliability of this modified bisection task. We hypothesized that humans will be able to determine intermediate durations significantly better than by chance, and that the modified temporal bisection task measuring the quantification of relative durations reaches a level of reliability acceptable for use in research.

3 Materials and Methods

3.1 Participants

Participants were recruited from Western University's SONA research pool. Twenty nine healthy human subjects (11 male, 18 female, N = 29) with normal hearing and of ages ranging from 18–46 participated in this study. One participant was excluded from the data due to failing attention checks. In total, data from 28 participants (n = 28) were analyzed. Consent was received from all participants prior to their participation in the study, and all procedures were approved by the Western University Non-Medical Research Ethics Board (REB approval code 106385).

3.2 Stimuli and Apparatus

The study was created using the software PsychoPy3 (version 2020.2.10) and was conducted as an online experiment hosted with the Pavlovia platform. Stimuli used in the study consisted of constant 500 hz tones with durations ranging from 500 ms to 1000 ms for reference and intermediate stimuli, or a 1500 hz tone of 1000 ms duration for attention check stimuli. All auditory stimuli were generated using the audiowrite function in MATLAB (version R2022b). Participants used a web browser on their personal computers to complete the study.

3.3 Procedure

Participants performed an identical group of tasks in two sessions, with the second session taking place one to seven days after the first. In each session, participants completed both the classic and modified temporal bisection task. These tasks were counterbalanced randomly, with half the participants completing the classic task first, and half completing the modified task first for both sessions. Each task consisted of three sequential phases: reference training, practice trials, and experimental blocks.

During reference training, participants were trained on two reference tone durations. Alternating tone intervals of a 500 ms short reference tone and a 1000 ms long reference tone were labelled and presented to participants for a total of five times for each tone. The reference training process was identical for the classic and modified temporal bisection tasks.

In the experimental trials, participants responded to a block consisting of 12 tones. This block of tones was repeated 12 times for a total of 144 presented stimuli per task. These tones were identical between the classic and modified tasks, and were composed of 11 test tones of 500 hz and an attention check tone of 1500 hz. The 11 test tones ranged in duration from the short reference of 500 ms to the long reference of 1000 ms in incrementing steps of 50 ms. Within each block, the 12 tones were presented in random order. After the presentation of each tone, participants responded based on their perception of the tone’s length relative to the reference durations. In the classic task, participants were instructed to respond by pressing a button on the keyboard based on whether they perceived the presented tone to be the short or long reference duration. If the higher-pitched attention check was presented, participants instead responded by pressing a different key. During the modified task, participants responded by using the mouse to drag a visual analog slider on-screen. The slider ranged from “short” to “long” to represent the reference durations, and participants were instructed to place the slider marker at or in between the reference durations based on their perception of the presented tone’s length. When presented the attention check, participants instead responded by placing the marker to the extreme right of the slider. After every two experimental blocks, participants were allotted a one-minute break. Between the classic and modified bisection tasks, participants were allotted a two-minute break.

Immediately prior to the experimental blocks, participants completed practice trials to familiarize themselves with the response process. In practice trials, an ordered set of stimuli were presented to participants. Instruction was given on how to correctly respond to each stimulus, and participants were then given the opportunity to respond. In the classic task, the presented practice tones consisted of the short and long reference tones each presented twice in an alternating manner. For the modified task, in addition to the reference tones, three intermediate tone durations of 600 ms, 750 ms, and 900 ms were presented to participants to familiarize responses to intermediate durations. After these practice tones, participants were trained on responding to the attention check, a higher-pitched 1500 hz tone. They were instructed to respond using a unique key for the classic task, and by placing the slider to the far right in the modified task.

3.4 Statistical Analysis for Classic and Modified Task Outcomes

A measure commonly studied from the classic temporal bisection is the bisection point, referring to the tone duration where an equal proportion of responses are either “short” or “long”. The bisection point was calculated using methods previously described by Wearden, (1991). For each tone duration, the proportion of “long” responses across all trials and participants were calculated. A least squares linear regression was performed on the four points encompassing the steepest slope when comparing stimulus durations and “long” response proportions using the lm function in R. The linear equation derived from this regression was then used to calculate the bisection point using 50% long responses as a parameter (Wearden, 1991). The computed bisection point was compared to previous literature stating that bisection approximates below the arithmetic mean of the two reference durations (Kopec and Brody, 2010).

To analyze differences between the accuracy scores from estimations made in the modified bisection task versus random responses, the estimation errors of participant responses were compared to a null distribution consisting of a randomized set of responses made from the collected data. The estimation error was obtained for each modified task response by calculating the absolute difference between the estimated ratio and the stimulus ratio (Lagacé-Cusiac et al., 2023). The estimated ratio was the ratio between the slider’s position and the slider length, while the stimulus ratio was the relative stimulus duration compared to reference durations (Lagacé-Cusiac et al., 2023). For each participant, the mean estimation error for all trials was calculated, and the set of mean estimation errors for all participants composed the experimental data set. To create the randomized data set, the slider responses were randomized within participants prior to calculating estimation errors to separate actual response versus true stimulus duration pairs. Means of estimation errors for each participant’s randomized data were calculated to create the null distribution, representing a data set simulating random responses from participants. An unpaired t-test ($\alpha=0.05$) was conducted using the t.test function in R to compare the actual distribution of mean participant estimation errors to the randomized null distribution of mean errors.

3.5 Statistical Analysis for Measures of Reliability

Internal-consistency reliability was estimated for both the classic and modified temporal bisection tasks using permutation-based split-half reliability (Parsons et al., 2019). For the classic task, participant responses were first randomly split into halves, balancing participant responses for each stimulus duration between halves. Within each split, the Weber ratio was determined for each participant as a measure of response accuracy in the classic task. The Weber ratio was calculated as the ratio between the difference limen and the participant's bisection point (Elvevåg et al., 2003). The difference limen was calculated as half the duration between where the proportion of "long" responses were 0.25 and 0.75 as determined by the linear regression line encompassing the bisection point (Wearden, 1991). A Pearson's correlation coefficient was then calculated to correlating accuracy scores between each half of the split. One thousand permutations of splits and correlations were performed, and the mean of the Pearson's r values was determined alongside the 95% confidence interval. To account for each split-half only having half the data of the complete classic bisection task, the Spearman-Brown prophecy formula was used on the mean correlation coefficient and confidence interval to estimate the internal-consistency reliability of this task (Parsons et al., 2019). For the modified task, the same balanced split halves were created, but the mean estimation error within each participant was used as the measure of response accuracy. Similarly, a Pearson's correlation was conducted between each of the 1000 permutations of split halves, and the mean Pearson's r and 95% confidence interval were subjected to the Spearman-Brown prophecy formula to estimate the internal-consistency reliability of the modified bisection task (Parsons et al., 2019).

To measure test-retest reliability for the classic and modified temporal bisection task, intraclass correlations were used. For the classic task, the Weber ratio from each participant in the first session of the experiment was correlated to their score in the second session. A mean-rating, absolute agreement, two-way mixed effects model intraclass correlation was conducted on this data using the R package irr version 0.84.1 to generate an intraclass correlation coefficient (ICC) and 95% confidence interval (Koo and Li, 2016; Gamer et al., 2019). A two-tailed F-test ($\alpha = 0.05$) was also conducted using the same R package to test for significance in the correlation (Gamer et al., 2019; McGraw and Wong, 1996). For the modified bisection task, a mean-rating, absolute agreement, two-way mixed effects model intraclass correlation was conducted on the mean estimation error for participants between the first and second session (Koo and Li, 2016). The ICC and 95% confidence interval were calculated, and a two-tailed F-test ($\alpha = 0.05$) was similarly conducted using the irr package (Gamer et al., 2019).

To determine relationships between measures of response accuracy in the classic and modified temporal bisection, convergent validity analyses were performed between the tasks. A Pearson's correlation was performed between participant Weber ratios for the classic task and mean estimation errors for the modified task using the cor.test function in r. The attenuation-correction formula was used on the resulting correlation coefficient to estimate the corrected correlation coefficient if each task measure had perfect reliability (Nicewander, 2018). The Spearman-Brown corrected internal-consistency reliability estimates for the classic and modified bisection task were used in the calculation of the attenuation-corrected correlation coefficient.

4 Results

4.1 Classic Bisection Task – Bisection Point

In the classic temporal bisection task, participants (n = 28) were trained on short and long reference tone durations before responding “short” or “long” to intermediate tones based on their perception of the tone’s duration. The proportion of “long” responses was calculated across all participants and trials for each stimulus duration (Fig. 1). The bisection point, obtained by performing a least squares linear regression on the four points with the steepest slope and finding the stimulus duration when the proportion of long responses reaches 0.5, was determined to be at 697.34 ms (Fig. 1) (Wearden, 1991).

4.2 Modified Bisection Task – Comparing True and Randomized Responses

For the modified temporal bisection task, participants (n = 28) were trained on short and long reference tone durations before responding to intermediate tones by placing them on a visual analog scale. The mean durations indicated by slider response during experimental trials were calculated for each stimulus duration (Fig. 2). Estimation errors were calculated as the absolute difference between estimated ratio on the slider and the stimulus ratio. An experimental distribution containing the mean estimation error for each participant was compared to a null distribution containing mean estimation errors for randomized slider responses within participants using an unpaired t-test (α = 0.05) (Fig. 3). The estimation error for true responses (M = 0.187, SD = 0.0516) was found to be significantly lower than for randomized responses (M = 0.369, SD = 0.0586); ($t(54) = -12.305$, $p < 0.0001$) (Fig. 3).

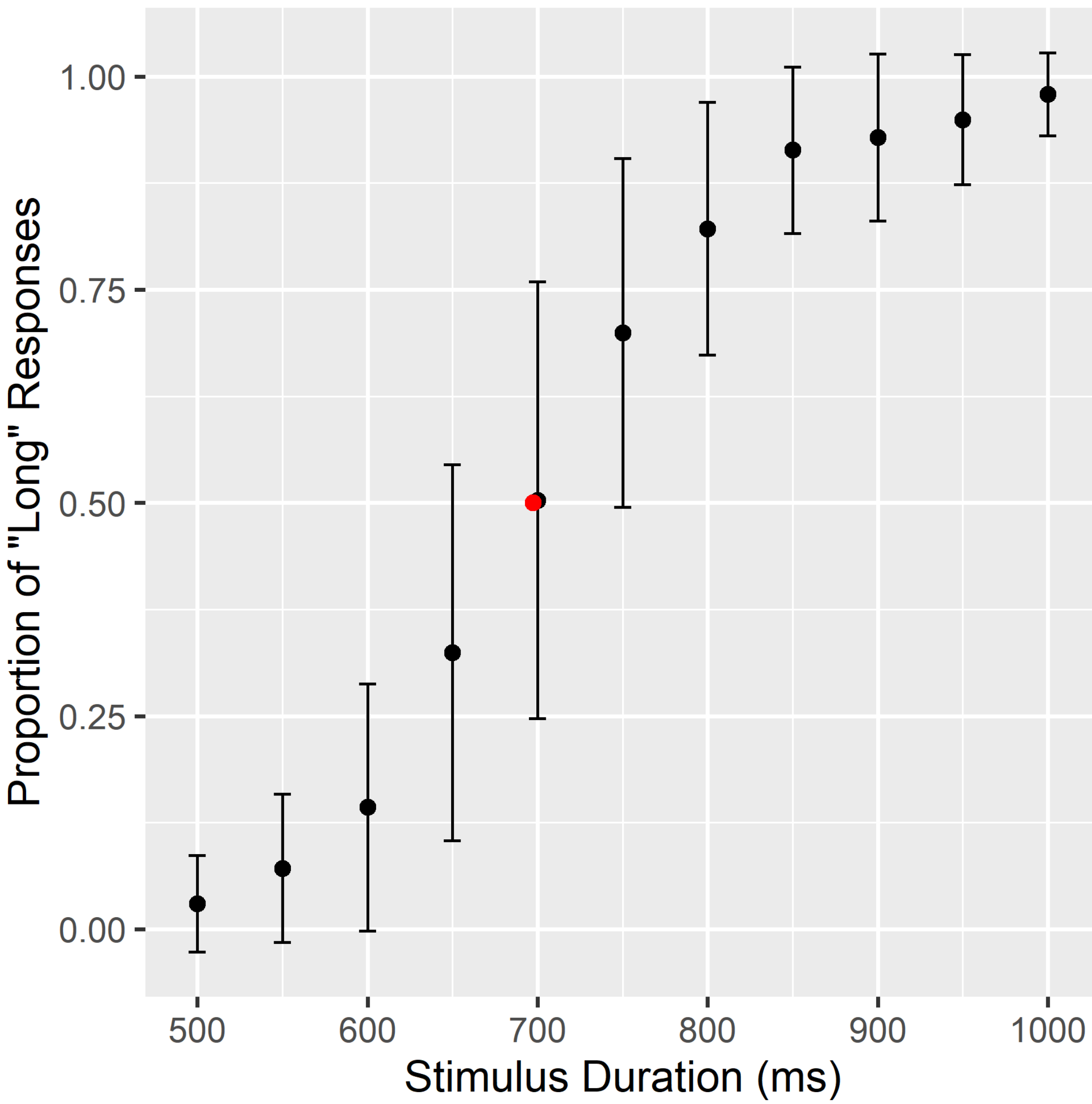

Figure 1: Proportion of “long” responses at presented stimuli durations (ms) for all participants (n=28) in a classic temporal bisection task. Participants were trained on reference tones of 500 ms (short) and 1000 ms (long). Data shown in black are the proportion of long responses across all participants ± standard deviations of proportions between participants. The bisection point (697.34 ms), where the proportion of “long” responses reached 0.5, is indicated in red.

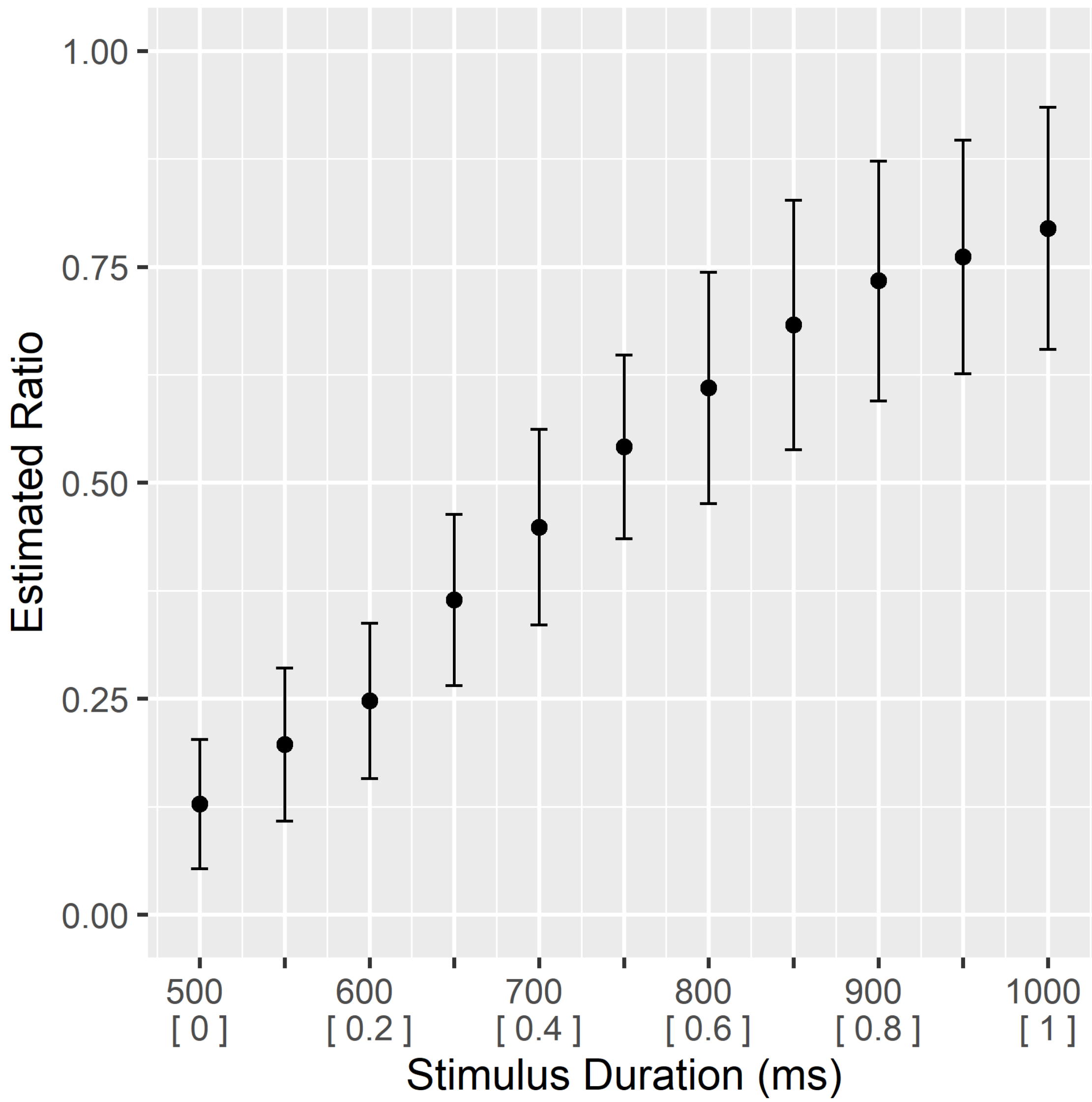

Figure 2: Mean estimated ratio on slider at presented stimuli durations (ms) for all participants (n=28) in a modified temporal bisection task. Numbers indicated in square brackets on x-axis indicate the stimulus ratio. Participants were trained on reference tones of 500 ms (short) and 1000 ms (long). Participants responded to intermediate stimuli on a continuous slider in the modified task. Data shown are the duration indicated by slider response (ms) across all participants ± the standard deviation for responses at each stimulus duration.

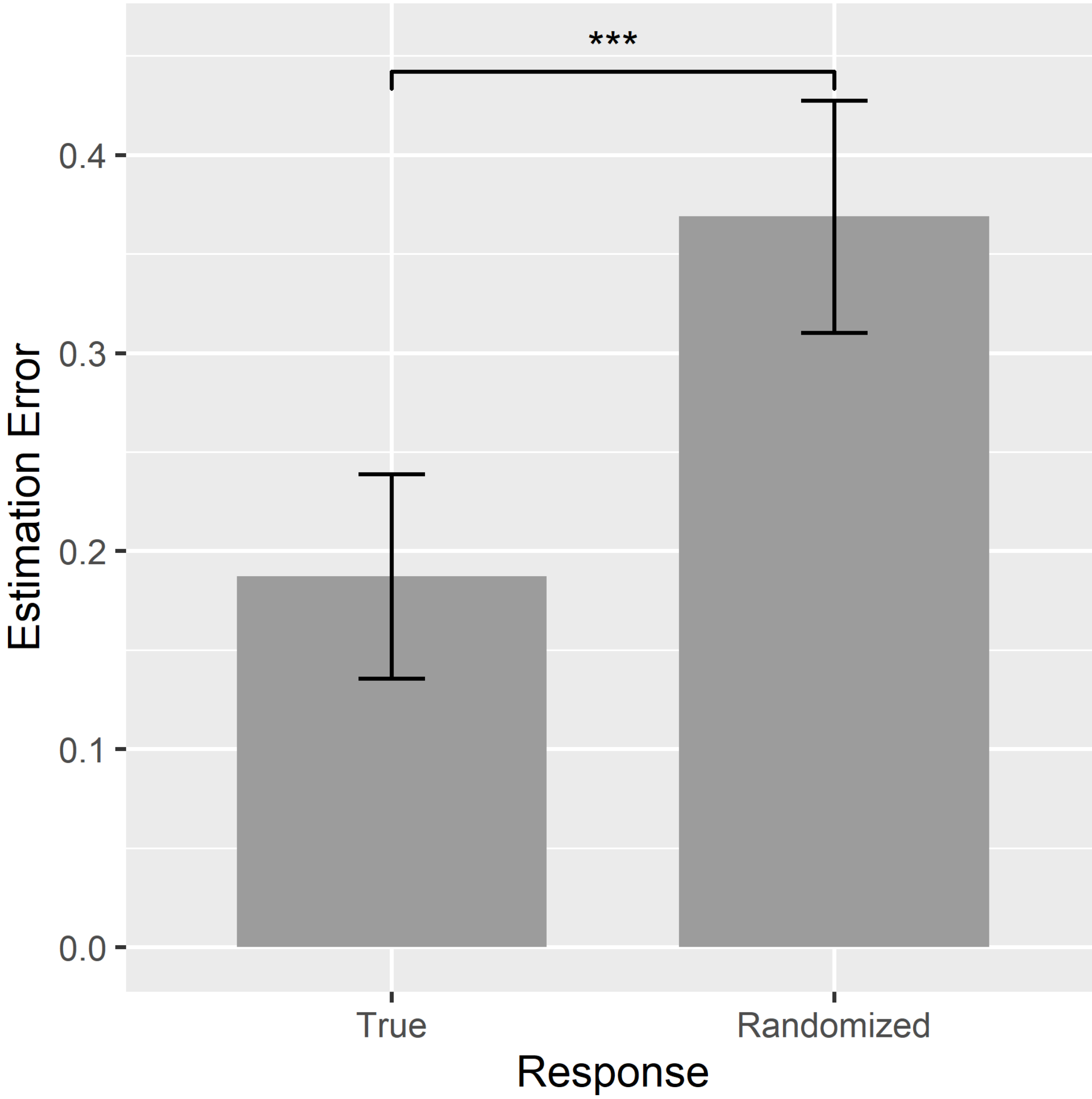

Figure 3: Mean estimation error across all presented stimulus durations and participants (n=28) for true participant responses versus randomized participant responses in a modified temporal bisection task. Participants were trained on reference tones of 500 ms (short) and 1000 ms (long). Participants responded to intermediate stimuli on a continuous slider in the modified task. Randomized responses were generated by randomizing slider responses within participants. Data shown are the mean estimation error ± standard deviation. Asterisks (***) between bars indicate a significant differences ($p < 0.0001$) following unpaired t-test ($\alpha=0.05$).

4.3 Internal-Consistency Reliability

Accuracy scores from the classic bisection task, measured by the Weber ratio, were calculated as half the difference between durations where proportions of “long” responses were 0.25 and 0.75, divided by the bisection point (Elvevåg et al., 2003). Internal-consistency reliability of the Weber ratio from the classic task was estimated using permutation-based split-half reliability (Parsons et al., 2019). One thousand permutations of random splits were generated, with data for each half balanced for participant responses to each stimulus duration. The Weber ratio was calculated for each participant (n = 28) within splits, and scores were correlated using a Pearson’s correlation between splits. The mean uncorrected correlation between all splits in the classic bisection task was positive (r = 0.370) with a 95% confidence interval of [0.361, 0.379]. When using the Spearman-Brown prophecy formula to predict split-half reliability across the whole test, the internal-consistency reliability of the classic task was estimated to be 0.540 with a 95% confidence interval of [0.531, 0.549].

Internal-consistency reliability of the estimation error from the modified bisection task was also estimated using permutation-based split-half reliability (Parsons et al., 2019). The mean estimation error of each participant (n = 28) within each split was calculated, and scores were correlated using a Pearson’s correlation between splits. The mean uncorrected correlation for all splits in the modified bisection task was positive (r = 0.931), with a 95% confidence interval of [0.929, 0.932]. When correcting the mean correlations using the Spearman-Brown prophecy formula to estimate split-half reliability across the whole test, the internal-consistency reliability of the modified task was estimated to be 0.964 with a 95% confidence interval of [0.963, 0.965].

4.4 Test-Retest Reliability

Participants ($n = 20$) completed the classic temporal bisection task for two sessions, with the second session taking place one to seven days after the first. Weber ratio scores from the first session were plotted against scores from the second session (Fig. 4) Intraclass correlations based on a mean-rating, absolute agreement, two-way mixed effects model were used to estimate the test-retest reliability for the Weber ratio in the classic task (Koo and Li, 2016). The intraclass correlation coefficient (ICC) for the test-retest reliability of the classic bisection task was measured as 0.506, with a 95% confidence interval of [-0.28, 0.806]. However, an F-test ($\alpha = 0.05$) conducted on intraclass correlation data showed that the correlation was not significant ($F(19,19.3) = 1.98$, $p = 0.071$) between scores in the first and second session for the classic bisection task. Noticing large outliers in the data, univariate analysis was performed on Weber ratios from both sessions to extract them (Fig. 4) (Kassambara, 2023). Sensitivity analysis was performed to examine the impact of these outliers. An intraclass correlation was performed on the same data with the two outliers removed, and the ICC was measured as -1.52, with a 95% confidence interval of ($-8.78 < ICC < 0.16$).

Participants ($n = 20$) also completed the modified temporal bisection in two sessions in the same arrangement as the classic task. Estimation error scores from the first session were plotted against scores from the second session (Fig. 5) An intraclass correlation based on a mean-rating, absolute agreement, two-way mixed effects model was conducted on the estimation error to estimate test-retest reliability in the modified task (Koo and Li, 2016). The ICC for the test-retest reliability of the modified task was measured as 0.902, with a 95% confidence interval of [0.757, 0.961]. An F-test ($\alpha =$

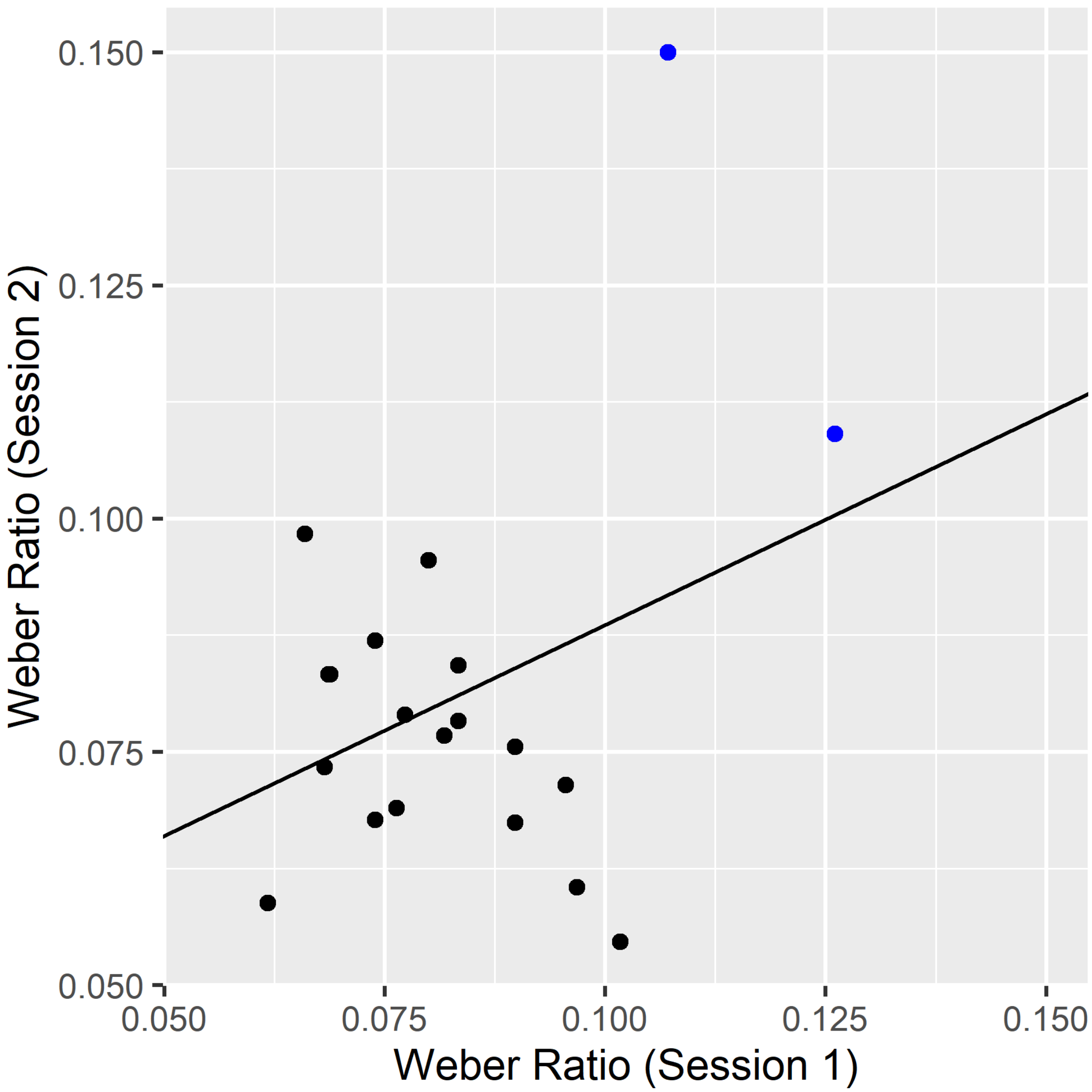

Figure 4: Weber ratio scores as a measure of classic bisection task accuracy between the first and second session, with the latter taking place 1-7 days after the first. Participants (n=20) were trained on reference tones of 500 ms (short) and 1000 ms (long). Data shown are the Weber ratio for each participant in both sessions. Data in blue are outliers as examined by univariate analysis of Weber ratios from either session. The linear regression between Weber ratios for the first and second session is indicated by the solid line.

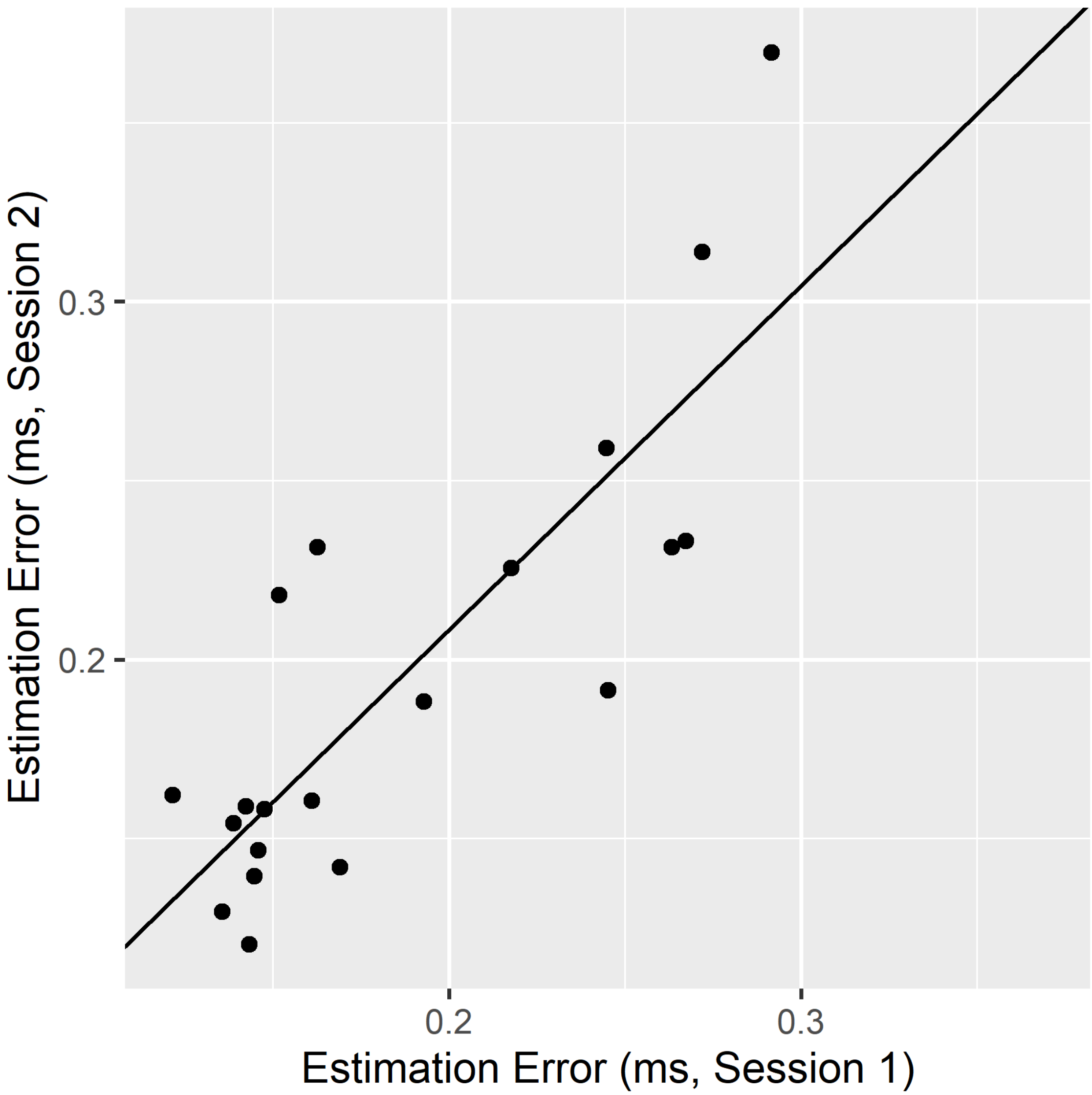

Figure 5: Estimation error as a measure of modified bisection task accuracy between the first and second session, with the latter taking place 1-7 days after the first. Participants (n=20) were trained on reference tones of 500 ms (short) and 1000 ms (long). Participants responded to intermediate stimuli on a continuous slider in the modified task. Data shown are the estimation error for each participant for each session. The linear regression between estimation errors for the first and second session is indicated by the solid line.

0.05) conducted on intraclass correlation indicated that the correlation in scores between sessions for the modified bisection task was significant ($F(19,20) = 10.3$, $p < 0.0001$)

4.5 Convergent Validity

After completing the classic and modified temporal bisection task, convergent validity analysis was conducted on accuracy scores from both tasks. Weber ratios for the classic task, and mean estimation errors for the modified task were analyzed with a Pearson's correlation using data from all participants ($n = 28$). A significant positive correlation ($r(26) = 0.396$, $p = 0.0369$) was found between classic and modified task accuracy scores with a 95% confidence interval of [0.0271, 0.670] (Fig. 6). The correlation coefficient between the Weber ratio and estimation error was corrected using the attenuation-correction formula to estimate the correlation given perfect reliability of either task (Parsons et al., 2019). Using the Spearman-Brown corrected internal-consistency reliabilities of 0.540 for the classic task, and 0.964 for the modified task in the attenuation-correction formula, the corrected correlation coefficient was calculated to be 0.549.

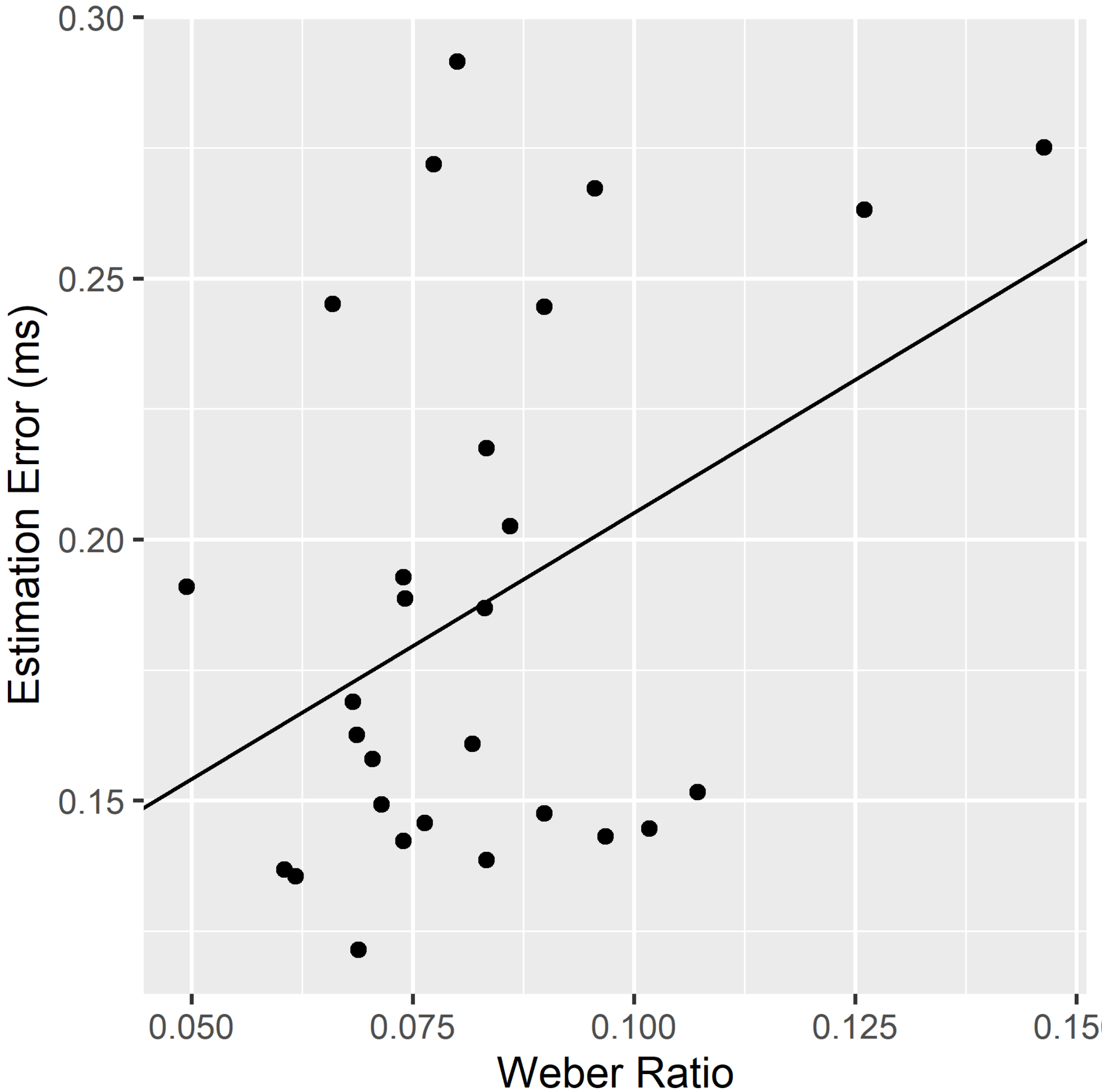

Figure 6: Comparisons between classic bisection task accuracy scores (Weber ratio) and modified bisection task accuracy scores (estimation error). Participants (n=28) were trained on reference tones of 500 ms (short) and 1000 ms (long). Participants responded to stimuli on a continuous slider for the modified task. Data shown are the Weber ratio and mean estimation error for each participant. The linear regression between the Weber ratio from classic bisection and the estimation error from the modified bisection is indicated by the solid line. A Pearson's correlation revealed a Pearson's r of 0.396.

5 Discussion

In the classic bisection task, the bisection point was determined to be at 697.34 ms (Fig. 1). The bisection point was slightly below the arithmetic mean between the two reference durations at 750 ms, which is in agreement with previous analyses of the bisection point (Kopec and Brody, 2010. The consistency of this measure with existing literature provides evidence of the sound implementation of the procedure, and provides support for the validity of the current study's classic task measures.

Examining results from the modified bisection task, true participant responses were significantly more accurate than randomized participant responses in terms of estimation error (Fig. 3). This supports our hypothesis that participants would be able to quantify relative durations in this task. The evidence that participants were able to achieve higher accuracy than random responses suggests that the intermediate durations presented as stimuli in the bisection task were likely perceived as intermediate durations, rather than simply as one of the reference durations.

As a measure of accuracy, the estimation error from the modified bisection task demonstrated excellent reliability across both internal-consistency and test-retest constructs. A Spearman-Brown corrected permutation based split-half correlation of 0.931 indicated excellent internal-consistency reliability, while an ICC of 0.902 indicated excellent test-retest reliability (Cicchetti, 1994). In contrast, the classic bisection task used the Weber ratio as a score of accuracy, which had a Spearman-Brown corrected correlation of 0.540 to indicate poor internal-consistency reliability (Cicchetti, 1994). In terms of test-retest reliability, an F-test conducted on intraclass correlation data indicated no significant correlation in Weber ratio scores between sessions. This demonstrates a lack of evidence for agreement between session scores (McGraw and Wong, 1996). However, the inconclusive results regarding the intraclass correlation for the classic task could be an effect of the small sample size used, especially given the large variability of the 95% confidence interval for the ICC of [-0.28, 0.806]. The sensitivity analysis determined that the removal of two outliers from the data shifted that ICC greatly, changing from 0.506 to -1.52. The high sensitivity of this result to the removal of these data indicate that the results seen may not have been robust, and could benefit from a larger sample size.

Results from analyzing convergent validity between accuracy scores in the classic and modified bisection task indicated a positive correlation of 0.396 (Fig. 6). This correlation is relatively weak, suggesting poor convergent validity between measures (Grobler and Joubert, 2018). However, the limited agreement between the two measures may be due to the imperfect reliability of the classic and modified bisection task. After considering the imperfect reliability of the measures by applying the attenuation-correction formula, the convergent validity between classic and modified task accuracy scores was estimated to be 0.549 (Nicewander, 2018). This moderate correlation provides some evidence of convergence, suggesting that the two tasks may be measuring the same underlying construct (Grobler and Joubert, 2018).

The excellent reliability of the modified temporal bisection task supports its use as a psychophysical measure. In the future, the modified task could be used to further explore the internal processes behind how humans perceive relative durations in time. Similarly to how the classic bisection task has been used to determine differences in temporal perception ability between clinical populations, the modified task could also be used to assess individual or group differences in subsecond temporal judgement accuracy (Liu et al., 2022; Nichelli et al., 1996; Elvevåg et al., 2003). The moderate corrected correlation validity between classic and modified task accuracy scores indicate that the modified task may be used in place of the classic task for determining accuracy in temporal judgements, with the advantage of being highly robust in reliability. However, in situations involving participants with relatively lower cognitive capacity, such as animal, pediatric, or geriatric testing, the classic bisection task may offer an advantage due to its relatively lower cognitive demand (Siegel and Church, 1984; Droit-Volet and Wearden, 2001; McCormack et al., 1999). In future applications of the classic bisection task, reliability analyses should be performed alongside its use. The poor reliability displayed by this task in the current study highlights the importance of assessing the consistency of this measure to ensure quality in results.

One of the limitations of the current study is the relatively small sample size. This limitation could be a contributing factor to the high sensitivity of results from the test-retest reliability of the classic task. A larger sample size would strengthen the generalizability and robustness of the findings. Obtaining a more comprehensive range of responses from the population for both temporal bisection tasks may also lead to a more conclusive evaluation of their convergent validity. In the current study, the convergent validity between the modified and classic bisection tasks was found to be inconclusive, ranging from poor to moderate. Collecting data from more participants could potentially capture a more robust relationship between classic and modified task accuracy scores. Additionally, participants in the current study were primarily composed of young adults because recruitment was done through Western University's SONA research pool. Considering a wider age range could improve the generalizability of the findings and further support the use of the modified bisection task. Conducting data analysis using an invariant framework and structural equation modeling could also improve the validity of the analyses. By correlating latent variables and controlling for measurement error, a more accurate understanding of the reliabilities and relationship between scores from each task could be obtained (Putnick and Bornstein, 2016).

In the future, conducting a larger-scale study including a wider variety of demographics to replicate these preliminary findings could provide further information about the robustness of the modified bisection task. Additionally, the human ability to quantify relative durations could be further studied using the modified bisection task. Comparing responses of this memory-dependent procedure to a comparable memory-independent procedure such as an empty auditory interval ratio estimation task could give insight on the role of memory in making relative temporal judgements (Lagacé-Cusiac et al., 2023).

Overall, the current study's aims were to determine if humans were able to quantify relative durations in a novel modification to the temporal bisection task, and to assess its reliability. Humans were significantly more accurate at determining intermediate durations in the modified task than random responses, showing that the modified task was able to capture the human ability to perceive relative durations. The modified bisection task also reached an excellent level of reliability in an internal-consistency and test-retest construct. It is based on these psychometric properties that the modified bisection task is recommended for future use to study internal processes of relative time perception and to assess differences in temporal perception ability.

6 Acknowledgements

I would like to express my thanks to Rebekka Lagacé-Cusiac, Dr. Jessica Grahn, and the Music and Neuroscience Lab for their continued support and assistance throughout the year. I would also like to acknowledge the Department of Physiology and Pharmacology at Western University for their guidance on this project. Finally, I would like to extend a thanks to our funding bodies James S. McDonnell Foundation, who provided the Scholar Award to Dr. Jessica Grahn (DOI:10.37717/220020403), and the Natural Sciences and Engineering Research Council of Canada, who supported Dr. Jessica Grahn through the Steacie Fellowship.

References

Allan LG, and Gibbon J (1991) Human bisection at the geometric mean. *Learning and Motivation* **22**:39–58.

Church RM, and Gibbon J (1982) Temporal generalization. *J Exp Psychol Anim Behav Process* **8**:165–186.

Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* **6**:284–290.

Droit-Volet S, and Wearden JH (2001) Temporal bisection in children. *Journal of Experimental Child Psychology* **80**:142–159.

Elvevåg B, McCormack T, Gilbert A, Brown GDA, Weinberger DR, and Goldberg TE (2003) Duration judgements in patients with schizophrenia. *Psychol Med* **33**:1249–1261.

Gamer M, Lemon J, and Singh IFP (2019) *irr: Various Coefficients of Interrater Reliability and Agreement*.

García-Pérez MA (2014) Does time ever fly or slow down? The difficult interpretation of psychophysical data on time perception. *Front Hum Neurosci* **8**.

Grobler A, and Joubert YT (2018) Psychological Capital: Convergent and discriminant validity of a reconfigured measure. *South African Journal of Economic and Management Sciences* **21**:14.

Kassambara A (2023) *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*.

Koo TK, and Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* **15**:155–163.

Kopec CD, and Brody CD (2010) Human performance on the temporal bisection task. *Brain Cogn* **74**:262–272.

Lagacé-Cusiac R, Tremblay PF, Ansari D, and Grahn JA (2023) Investigating the relationships between temporal and spatial ratio estimation and magnitude discrimination using structural equation modeling: Evidence for a common ratio processing system. *J Exp Psychol Hum Percept Perform* **49**:108–128.

Lindbergh CA, and Kieffaber PD (2013) The neural correlates of temporal judgments in the duration bisection task. *Neuropsychologia* **51**:191–196.

Liu P, Guo H, Ma R, Liu S, Wang X, Zhao K, Tan Y, Tan S, Yang F, and Wang Z (2022) Identifying the difference in time perception between major depressive disorder and bipolar depression through a temporal bisection task. *PLoS One* **17**:e0277076, Public Library of Science.

Matthews WJ, and Meck WH (2014) Time perception: the bad news and the good. *Wiley Interdiscip Rev Cogn Sci* **5**:429–446.

McCormack T, Brown GDA, Maylor EA, Darby RJ, and Green D (1999) Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology* **35**:1143–1155.

McGraw KO, and Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* **1**:30–46.

Nakajima Y (1987) A model of empty duration perception. *Perception* **16**:485–520, SAGE Publications Ltd STM.

Nicewander WA (2018) Modifying Spearman's attenuation equation to yield partial corrections for measurement error—with application to sample size calculations. *Educ Psychol Meas* **78**:70–79.

Nichelli P, Alway D, and Grafman J (1996) Perceptual timing in cerebellar degeneration. *Neuropsychologia* **34**:863–871.

Parsons S, Kruijt A-W, and Fox E (2019) Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Adv Methods Pract Psychol Sci* **2**:378–395, SAGE Publications Inc.

Putnick DL, and Bornstein MH (2016) Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Dev Rev* **41**:71–90.

Read JCA (2015) The place of human psychophysics in modern neuroscience. *Neuroscience* **296**:116–129.

Siegel SF, and Church RM (1984) The decision rule in temporal bisection. *Ann NY Acad Sci* **423**:643–645.

Wearden JH (1991) Human Performance on an Analogue of an Interval Bisection Task. *Q J Exp Psychol B* **43**:59–81, SAGE Publications.

Wearden JH, and Ferrara A (1995) Stimulus Spacing Effects in Temporal Bisection by Humans. *Q J Exp Psychol B* **48**:289–310, Routledge.

Wittmann M (2009) The inner experience of time. *Philos Trans R Soc Lond B Biol Sci* **364**:1955–1967.