Federal University of ABC

Postgraduate Program in Neuroscience and Cognition

Adaptation of the Temporal Bisection Task for Remote Collection: Development, Analysis and Validation

Marcel Valério de Arruda[1,2]
Bruna Verzili Gallo[1]
Flavio Herrmann[1]
Soraia Fernandes das Neves Glisoi[1]
Ruth Ferreira Galduróz[1]
Marcelo B. Reyes[1]

1 - UFABC - Center for Mathematics, Computation and Cognition / Postgraduate Program in Neuroscience and Cognition
2 - Corresponding Author: marcel.a@ufabc.edu.br

ABSTRACT: The COVID-19 pandemic posed a significant impediment to experimental research, leading several researchers to adapt psychophysical data acquisition. With the development and proliferation of information technology, paper-based organizational processes have gradually begun to be replaced by computer-based equivalents. This study describes how we developed, analyzed data, and validated the temporal bisection task in a remote data acquisition scenario. We implemented the data acquisition using the open science software OpenSesame, in conjunction with the JATOS platform. We described in detail all the steps to use our codes, which we made available for reuse. We acquired data from 28 participants using the remote acquisition system and compared them with data obtained in person (from Penney et al., 2000). Our remote data showed compatible results with live experiments, suggesting that the modality of data acquisition (remote or live) does not influence the results.

Keywords: Temporal Bisection Task; Human Timing; Interval Timing; Time Perception; Bisection Point; Weber's Ratio; Differential Threshold; Validation; Remote Data Acquisition.

Public Significance Statement - Results of this study revealed that the modality of data acquisition (remote or live) does not influence the results. The results of these experiments add to the growing number of studies throughout the COVID-19 pandemic, encouraging remote data collection.

Word Count: 4403

São Paulo, Brazil

2022

1. INTRODUCTION

With the development and proliferation of information technology, paper-based neuropsychological assessments have gradually begun to be replaced by computer-based equivalents (Stone et al., 2015). Psychometric tests are no exception. Particularly, experiments that require on-site responses were unfeasible due to the pandemic of the SARS-CoV-2, which urged experimenters to adapt experiments to remote environments. Such adaptation may provide two advantages: a) improving the measurement accuracy by removing human error in calculating standardized scores, and b) accelerating the scoring process, potentially allowing to score thousands of responses in seconds – something that might take many hours or days with paper-based tests. Finally, virtual psychometrics tests may provide researchers with much larger data samples - from a broader range of cultural and demographic backgrounds - than was previously practicable (Hinton et al., 2016).

The field of timing and time perception has decades of accumulated information on the use of psychometric experiments aiming to unravel functional and neural mechanisms underlying time processing (Thoenes et al., 2017; Matthews et al., 2016; van Rijn et al., 2014; Kopec et al., 2010). One of the commonly used tasks to study timing is the temporal bisection task (Church and Deluty, 1977; Stubbs, 1968; Catania, 1970) to study temporal discrimination in rats. Since then, many researchers adapted the behavioral procedure to humans (Wearden, Allan, and Gibbon, 1991; Kopec et al., 2010, Jozefowiez, 2018).

In the standard Temporal Bisection task, participants learn to classify two reference stimulus durations as "Short" and "Long." For example, participants learn to press one key after a 0.5 s (Short) stimulus and another after a 2.0 s (Long) stimulus. Afterward, they must judge whether the duration of a probe stimulus seems closer to the Short or the Long reference stimulus, pressing the corresponding key. The responses allow researchers to identify intervals that participants report as short and long with the same probability, called the bisection point. The bisection point (BP) represents the duration that the subject classifies 50% of the trials as "short" and 50% as "long." The task also identifies how steep is the transition of participants from judging short and long stimuli around the bisection point through a quantity called the just noticeable difference (JND). The differential threshold (DT) is the minimum amount by which changes in stimulus intensity produce a noticeable variation in sensory experience (Levy et al., 2015).

This study describes how we developed, analyzed data, and validated the temporal bisection task in a remote data acquisition scenario. Each participant performed the task on his personal computer, which sent the responses via the JATOS platform. Our results show that the data is consistent with results from the literature, suggesting that the new acquisition methodology is trustworthy. Finally, we made our code available to the scientific community for reuse, adaptation, and expansion.

2. METHODOLOGY

2.1. *OpenSesame* Software & JATOS Platform

The temporal bisection task was developed using the open science software *OpenSesame*, in conjunction with the JATOS (*Just Another Tool for Online Studies*) platform that helps to set up and run online studies on their server (Lange et al., 2015). The availability of experiments on JATOS - developed in *OpenSesame* - is made possible by the *OSWeb* tool (Supplementary Fig. 1) (Mathôt et al., 2012;

Mathot & March, 2021). This tool allows to run the experiments in a web browser, export the developed experiments (program code) to the JATOS platform and convert the *.txt* files received from the participants by the JATOS platform into *.csv* (Supplementary Fig. 2), making it straightforward to load, share, and analyze data.

2.2. DEVELOPMENT: Temporal Bisection Task in the *OpenSesame*

When starting an experiment in *OpenSesame* it is necessary to open a sequence ( ) to instruct the program to perform a series of operations in a specific order. Next, we configured the two introductory pages using *sketchpad* ( ) (Supplementary Fig. 3).

The TB task itself was divided into three steps, each requiring a loop ( ) with a condition table holding the temporal specifications of the trial stimuli and responses. Two keyboard keys, 'Q' on the left side and 'P' on the right side, defined the correct choices following the Short and Long stimuli, respectively.

1st Phase: Setting the short and long sample stimuli required two loops. In *loop_short*, a 500 ms Short sample (TS) is presented five times, and the participant's *'short'* response was requested in all five trials (Supplementary Fig. 6). In *loop_long*, a 2000 ms Long sample (TL) is presented five times, and the participant's '*long'* response is requested on all five trials (Supplementary Fig. 7).

The templates presented in Supplementary Figures 6 and 7 were used throughout the experiment. First, a loop was created with the table required to execute each training trial. Then, the sequence of trial events was defined by three pages, each constructed with the sketchpad. The first page defined the 500 ms fixation point (Supplementary Fig. 8). The second page defined the sample stimulus, where it was presented in the display (Supplementary Fig. 9) and for how long (the stimulus duration was extracted from the table defined in the loop, column “duration”). Finally, the third page defined the response waiting period (Supplementary Fig. 10). Setting the page's Duration to 0 (zero) forces the experiment to remain on the page until the participant responds by pressing the correct key. That explains why in the 1st step (training phase) the top of the page displayed the instructions "Press 'Q' for the short stimulus" and "Press 'P’ for the long stimulus".

Right after the waiting page, the routine requires a response from the participant. This procedure was implemented by creating a *keyboard_response (*) (Supplementary Fig. 11). The response page specified the correct trial response (if any) using the condition table defining the trials, column [answer_correct] (Supplementary Figure 4). It also specified the allowed trial responses. To inform the participant that his response had been computed, a *sampler (*) was added (Supplementary Fig. 12), which plays a sound after each response.

At the end of each sequence, the participant's response is saved using the *logger (*), and then, according to what was selected in the *loop* the sequence, restarted or ended. When the sequence is finished, a *feedback* page ( ) should be created so that the participant understands that the sequence has ended, and prints instructions to the next sequence (Supplementary Fig. 13).

2nd Phase: Responses: Evaluation and learning. The complete sequence is shown in Supplementary Figure 14. The participants are presented with the short (500 ms) and long (2000 ms) stimuli in random order and, after each one, they press one of the two allowed keys, 'Q' for "Short" or 'P' for "Long". The phase lasts until the participant obtains ten correct trials.

The loop for the second stage uses a condition table similar to that shown in

Supplementary Fig. 4, but with two instead of one loop terminating conditions (see Supplementary Figure 15). In addition to the number of repeats (set to 100 in this case), the loop also ends if the total number of correct responses reaches 10. By default, OpenSesame stores the number of correct responses automatically.

The other components of the sequence are equivalent to those that were presented in the 1st phase. Introduced with the use of the *sketchpad*, the fixation point, the stimulus, and the waiting page are presented. As well as the sampler and the logger. There's a difference in the *keyboard_response*, where in the 1st phase only one response was allowed per sample stimulus ('Q' for "Short" and 'P' for "Long"). Now, both are allowed.

3rd Phase: Classification of intermediate probe stimuli. In this phase, seven different durations of the stimulus were presented. The intervals were 500, 630, 793.5, 1000, 1260, 1587, and 2000 milliseconds. Participants are instructed to respond to which classification, 'short' or 'long', the stimuli were more similar.

For the third phase, where the participants classify the intermediate probe stimuli as 'short' or 'long', the saved variables are redefined using *reset_feedback (* ), this is done as this is the *loop* of interest in this experiment, i.e., the loop that contains the information for the analysis regarding the bisection point, the DT and the WR. The other components of the sequence are equivalent to those presented in the 2nd phase. The fixation point, the stimulus, and the waiting pages, as well as the keyboard_response, the sampler, and the logger remain the same as before. The loop, however, differed: Its condition table included not only the two reference stimuli, but five additional stimuli with intermediate durations (Supplementary Figure 17). Moreover, the loop presented the 7 stimuli in random order, repeating itself for 10 times (see Supplementary Figure 18). Hence, phase three comprised a total of 70 trials. Participants received no feedback.

At the end of the experiment, the program saves the data in a ".csv" (i. e., comma separated variable) extension file that is used for analyses.

2.3. ANALYSIS

The results of a temporal bisection procedure are usually evaluated by computing the probability $P_\ell(T)$ of responding 'long' as a function of the sample duration T. The data is modeled with a sigmoid function (Fig. 1). From the sigmoid that best fits the results, we obtain two parameters: the BP and the DT. The Weber's ratio can be obtained by the ratio between these two parameters.

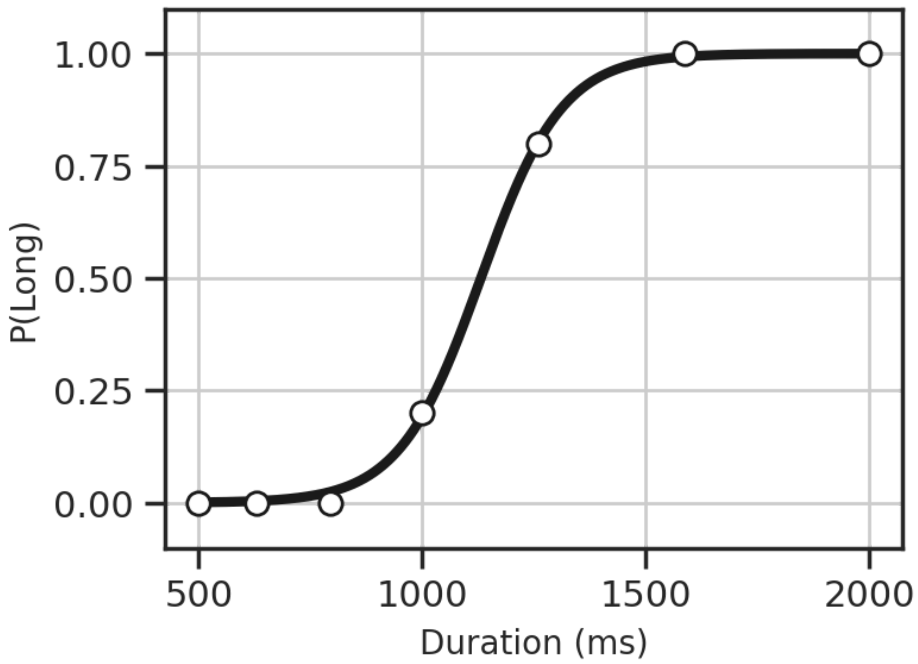


Figure 1 - Example of a psychometric function. The probability of responding long P(Long) obtained from a participant for each probe duration was shown in data points (open circles). The continuous line represents the model fit to the data point.

The bisection point (BP) is defined as the duration that the subject classifies 50% of the trials as 'short' and 50% as 'long'. The BP is the point of subject equality (PSE) because it identifies the duration that subjectively is equally distant from the

Short and Long reference durations.

An important quantity captured by the temporal bisection task is the slope of the function at the BP, which relates to the subject's sensitivity to variation in the duration of the sample stimulus (Fig. 2). The higher the sensitivity to the time dimension, the steeper the psychometric function at the BP (Kopec et al.,2010; Castro et al., 2013).

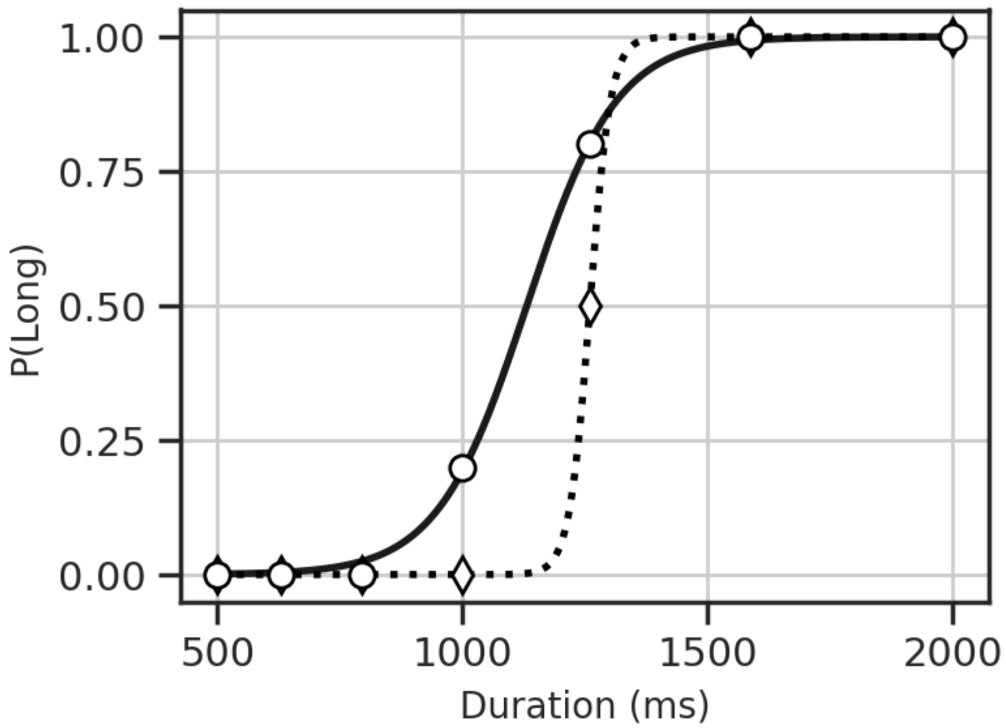


Figure 2 - Example of a low slope psychometric function (continuous line) and a high slope psychometric function (dotted line).

We estimated the slope of a specific psychometric function using the points T1=$T_{(P\ell=0,25)}$ and T2=$T_{(P\ell=0,75)}$, the durations with a 25% and 75% probability of 'long' responses, respectively. These points (T1 and T2) were obtained through the algorithm available in the Open Practices Statement. The differential threshold is defined as DT = T2-T1 divided by 2. The subject's sensitivity to the duration of the stimulus, the slope of the function in the PSE, and the DT are closely related, with sensitivity being directly proportional to the slope and inversely proportional to the DT.

The Weber's Ratio (WR) or Weber's Fraction, is defined according to the following equation: (T2 - T1) / BP. The difference (T2 - T1) is called the just noticeable difference (JND), which represents the smallest change in the stimulus that produces a substantial behavior change. A subject with a high degree of discriminability would produce a psychometric curve that appears very staggered, resulting in a low WR, while another subject with lower discriminability would produce a more gradual psychometric function, resulting in a higher WR (Kopec et al., 2010).

2.4. VALIDATION: Comparison Between Live & Remote Collection

Kopec and collaborators (2010) conducted a meta-analysis of human performance on the TB task. They collected data from 18 independent studies executed over the past two decades. The studies reported 148 experiments and involved approximately 1020 individuals discriminating stimulus durations on more than 300 thousand trials. Short reference durations (TS) ranged from 50 ms to 8 seconds, and long (TL) ranged from 200 ms to 32 seconds; the ratio TL/TS ranged from 1.2 to 19. In most studies, probe test durations were either linearly or logarithmically spaced. The stimuli were either visual - presented on the computer screen - or auditory. The authors combined all the results into a single data set (Kopec et al., 2010).

Among the 148 experiments evaluated, one experiment from the study by Penney et al. (2000) was similar to ours: they used subjects closer in age (adults and elderly), the same ratio TL / TS = 4, visual stimuli, and logarithmically spaced probe intervals.

We compared the live and remote data using 39 participants, 28 with remote data collection (16 adults and 12 elderly women), and 11 with live data acquisition (from Penney et al., 2000). Five participants were excluded because of poor performance on the TB task, and one participant was excluded because of a score above 10 on the Geriatric Depressive Symptom Scale (GDS - 30).

The live experiment assessed 23 naive undergraduate students from Columbia University (Live group) with three pairs of short and long anchor points (TS-TL): 3s - 6s, 4s - 12s, and 2s - 8s. All participants experienced all three duration ranges, but only 11 participants received visual stimulus sessions. For the validation described here, we only considered the results of the 2s - 8s test because the ratio TL / TS = 4 was identical to the one used in the remote collection. In the 2s - 8s test, the durations were 2.00, 2.52, 3.18, 4.00, 5.04, 6.35, and 8.00s. Each session consisted of 10 training trials and 100 test trials (Penney et al., 2000).

For the remote data collection, participants were asked to sit on a chair, 50 cm away from the computer monitor, and with headphones to minimize external noise. They were asked to access a Web page that displayed the JATOS platform and ran the OpenSesame program. The first page displayed was the welcome page with the initial instructions (see Supplementary Figure 19). On the keyboard of their personal computer, the participants were instructed to respond with their left hand, the 'Q' key whenever the desired response was 'short', and, with their right hand, the 'P' key when 'long'.

Initially, the participants went through a task-familiarization period (training phase), where they were presented with standard time durations of 500 ms and 2000 ms (TS and TL, respectively). Then a series of intermediate time durations ranging from 0.5 to 2 seconds (500, 630, 793.5, 1000, 1260, 1587, and 2000 milliseconds), where participants had to respond to which of the references (TS or TL) the presented interval was more similar (see section 2.2).

There were two groups for the remote task, one with 16 adults (24.43 ± 1.96 years old), 4 males and 12 females (group Adult), and the other with 12 elderly women

(62.91 ± 1.06 years old), the group Elderly.

In the remote assessment, participants were neuropsychiatricly assessed with the Geriatric Depressive Symptom Scale (GDS - 30) and with the Mini-Mental State Examination (MMSE) described below.

Geriatric Scale of Depressive Symptoms (GDS - 30): This scale consists of 30 questions with yes or no answers, aiming at quantifying depression signs and symptoms. The final score corresponded to the sum of the answers, with higher scores characterizing more intense depressive symptoms (Yesavage et al., 1982).

The 30-item GDS has a sensitivity of 84% and a specificity of 95% (Roman & Callen, 2008). Unlike other depression scales, the GDS does not include questions related to somatic complaints that could cause false negatives, since these complaints can be ambiguous. The 30-item GDS has remained unchanged since 1983 and has become the instrument most widely used by researchers and clinicians to diagnose depression (Roman & Callen, 2008; Edwards et al., 2004).

Mini-Mental State Examination (MMSE): This is an instrument with questions of seven categories, each evaluating specific cognitive functions: time orientation, location orientation, memory, attention and calculation, recall, language, and constructive visual ability. The MMSE score ranges from 0 to 30, with lower scores suggesting cognitive deficits (Folstein, 1975). Since the MMSE is influenced by education, reference values have been proposed to distinguish subjects with possible cognitive deficits. Brucki et al. (2003) analyzed a Brazilian sample, validating the scale in Brazil, and suggested the following values for studies in our country: for illiterates, 20 points; from 1 to 4 years of schooling, 25; from 5 to 8 years, 26.5; from 9 to 11 years, 28; and for individuals with more than 11 years of schooling, 29 points.

Neuropsychiatric assessments served as an exclusion criterion, cutting out participants who scored above 10 on the GDS-30 and below 23 on the MMSE.

The statistical analysis was conducted using statistical software *Jamovi* v1.6 (The Jamovi Project, 2021) and *Statistica* v14.0.0 (TIBCO Statistica, 2020) with a 95% confidence level. We applied the mean difference method, a standard statistical method that in this study, estimates the amount by which the remote collection changes the result on average compared to the live collection (Andrade, 2020). The mean difference method was conducted via the MAJOR module from *Jamovi* (Viechtbauer, 2010; Lakens, 2017). We compared two statistical analyses, with different software and researchers, and the results did not differ.

2.5. Transparency and Openness Statement

Regarding our compliance with the Transparency and Openness Promotion (TOP) guidelines. The data and materials for all experiments are available at <https://github.com/m-v-arruda/Temporal-Bisection-Task>, at the Supplementary Materials section and in the APPENDIX I. And none of the experiments were pre registered. The approved project has CAAE number (Certificate of Presentation of Ethical Appreciation): 42357021.0.0000.5594 and opinion number: 4.618.029 issued by CEP/UFABC (UFABC Research Ethics Committee).

3. RESULTS

As the short and long reference intervals are different in the remote and live collection studies, we used the arithmetic (AM) and geometric (GM) averages of the reference times to normalize the bisection points of the groups, thus enabling a comparison between the studies (Kopec et al., 2010). In remote collection the reference intervals were TS = 500 ms and TL = 2 s, while in live collection TS = 2 s and TL = 8 s (Table I).

Table I - Temporal bisection task conditions.

Mode	TS	TL	AM	GM	Proportion (TL/TS)
Remote	0.50	2.00	1.25	1.00	4
Live	2.00	8.00	5.00	4.00	4

Unit: Seconds; *AM* - Arithmetic Mean; *GM* - Geometric Mean; *TS* - Time Short; *TL* - Time Long.

The arithmetic (AM) and geometric mean (GM) of the reference durations are defined as:

$$AM = \frac{(TL + TS)}{2}$$

and

$$GM = \sqrt{(TL \times TS)}\,.$$

The bisection points obtained from data fit were divided by the arithmetic mean (BP/AM) and the geometric mean (BP/GM).

When comparing the results of the measures extracted from the temporal bisection task between the Elderly group (remote) and the Live group, significant differences were observed in the BP and DT measures ($t(21) = -19.67$, $p < .0001$; $t(21) = -5.31$, $p < .0001$, respectively) (two-tailed, independent samples, t-test). Such difference was expected since the TS and TL varied greatly in these groups. However, no significant differences were observed in the normalized measures, BP/AM and BP/GM ($t(21) = -1.29$, $p = 0.210$; $t(21) = -1.31$, $p = 0.206$, respectively) or in WR ($t(21) = 0.98$, $p = 0.3364$) (Table II).

Table II - Mean differences of temporal bisection task (TB) measures between studies (Elderly).

TB Measures	Mode	x̄ (SD)	Mean Difference	p-value
Bisection Point	Remote	1.002 (0.157)	-3.328	< .0001*
	Live	4.330 (0.564)		
BP/AM	Remote	0.802 (0.125)	-0.064	0.2098
	Live	0.866 (0.113)		
BP/GM	Remote	1.002 (0.157)	-0.081	0.2055
	Live	1.083 (0.141)		
Differential Threshold	Remote	0.174 (0.070)	-0.516	< .0001*

	Live	0.690 (0.330)		
Weber’s Ratio	Remote	0.348 (0.135)	0.188	0.3364
	Live	0.160 (0.663)		

AM - Arithmetic Mean; *BP* - Bisection Point; *GM* - Geometric Mean; *SD* - Standard Deviation; $\bar{x}$ - Mean.
*For the Bisection Point and the Differential Threshold there is a significant difference.

We compared the normalized measures (BP/GM, BP/AM, and WR) extracted from the TB task between the Elderly and the Live group, computing the respective effect sizes (Fig. 3). An insignificant effect size ($d < 0.19$) is observed for measures of the BP normalized by the AM and GM and for the WR, in addition it can be seen that the 95% confidence interval (CI) crosses the null effect line ($d = 0.0$) for these measures, where there are no differences between the groups (Cohen, 1988; Rosenthal, 1996).

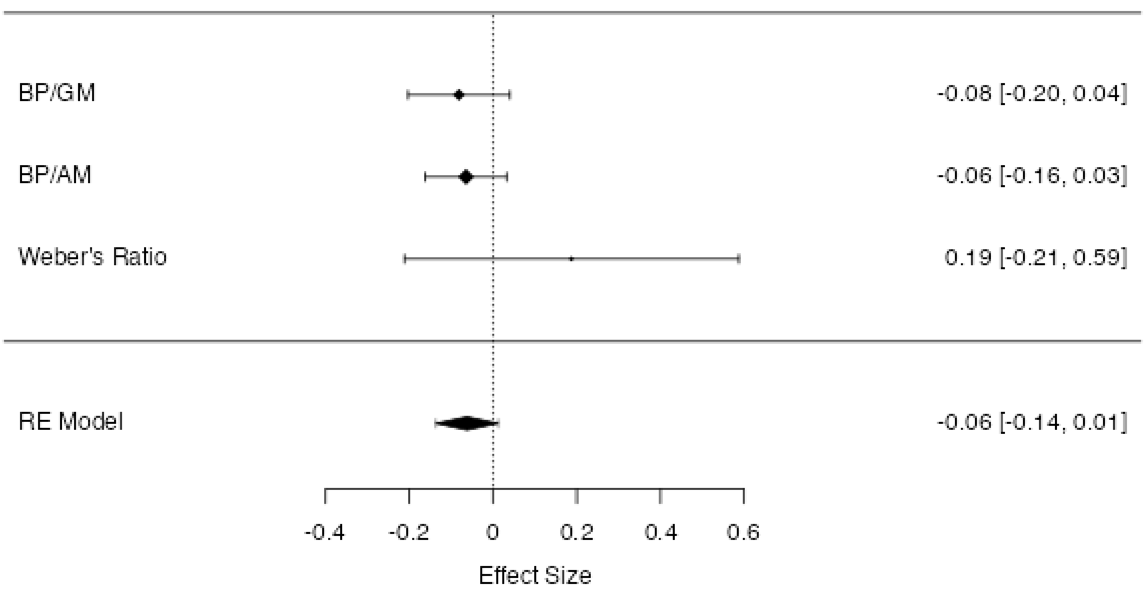


Figure 3 - Forest plot of Cohen's d effect sizes and corresponding 95 % confidence intervals for the elderly group.

The results from the Adult and the Live group, also differed in the BP and DT, as expected ($t(25) = -21.09$, $p < .0001$; $t(25) = -4.59$, $p = 0.001$, respectively). No significant differences were observed in BP/AM ($t(25) = -0.43$, $p = 0.6676$), in BP/GM ($t(25) = -0.43$, $p = 0.6684$) or in WR ($t(25) = 1.17$, $p = 0.2536$), as described in Table III.

Table III - Mean differences of temporal bisection task (TB) measures between studies (Adult).

TB Measures	Mode	x̄ (SD)	Mean Difference	p-value
Bisection Point	Remote	1.049 (0.226)	-3.281	< .0001*
	Live	4.330 (0.564)		
BP/AM	Remote	0.839 (0.180)	-0.027	0.6676
	Live	0.866 (0.113)		
BP/GM	Remote	1.049 (0.226)	-0.034	0.6684
	Live	1.083 (0.141)		
Differential Threshold	Remote	0.211 (0.213)	-0.479	0.001*
	Live	0.690 (0.330)		
Weber's Ratio	Remote	0.445 (0.594)	0.285	0.2536
	Live	0.160 (0.663)		

AM - Arithmetic Mean; *BP* - Bisection Point; *GM* - Geometric Mean; *SD* - Standard Deviation; $\bar{x}$ - Mean.
*For the Bisection Point and the Differential Threshold there is a significant difference.

Figure 4 shows the effect sizes for each of the comparisons made of the normalized measurements extracted from the TB task between the Adult group and the Live group.

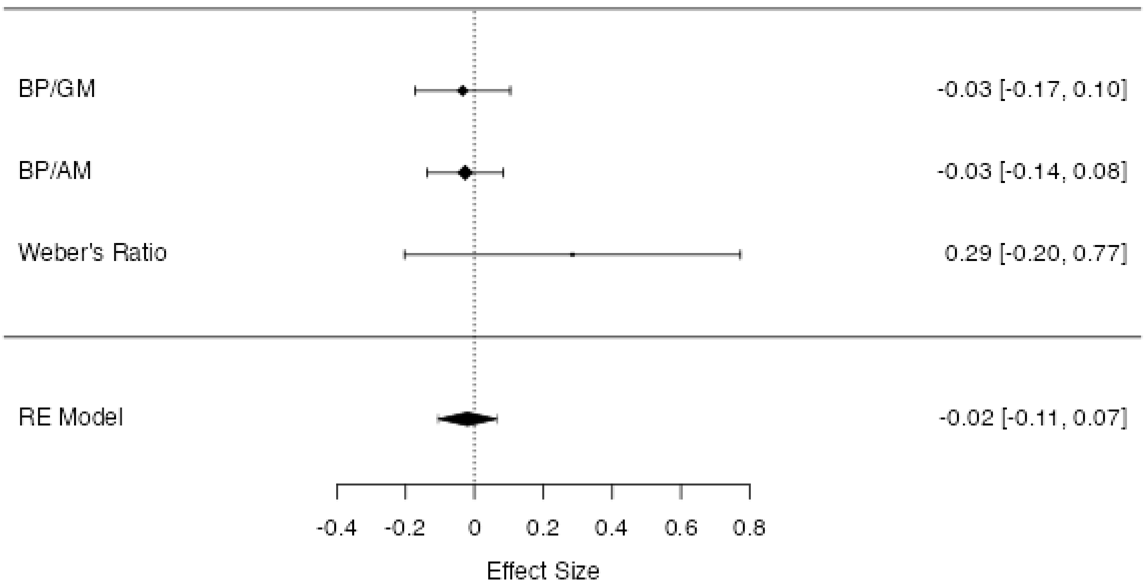


Figure 4 - Forest plot of Cohen's d effect sizes and corresponding 95 % confidence intervals for the adult group.

An insignificant effect size (d < 0.19) is observed for measures of the BP normalized by the AM and GM means and a small effect size (0.20 < d < 0.49) for the WR, in addition, the 95% CI crosses the null effect line (d = 0.0) for these measures, where there are no differences between the groups (Cohen, 1988; Rosenthal, 1996).

4. DISCUSSION

Acquiring good quality data online opens significant possibilities to reach wider population groups. Several studies intending to evaluate psychometric parameters use locally available participants. For example, researchers frequently recruit undergraduate students from the university where the study is conducted. Such idiosyncrasy in data samples may bias and mask results due to the particular characteristics of such a sample. For example, in a psychology university, students may have some knowledge about psychometric tests that can bias the results compared with those obtained with a more diverse population. Online (remote) data acquisition helps overcome these limitations, enabling researchers to increase sample diversity. It can also increase the sample sizes resulting in more solid and reproducible results.

Here we paved the way to acquire data online in a well-described interval timing procedure, the temporal bisection (TB) task. Since Church and Deluty (1977) systematic study of temporal bisection, this task has been widely used by researchers interested in time perception and its disruptions in patients afflicted with social anxiety (Jusyte et al., 2015), depression (e.g., Gil & Droit-Volet, 2009), or Parkinson's disease (e.g., Terao et al., 2021), for example.

Even though online questionnaires are ubiquitous, online psychometric data acquisition faces complex technical challenges such as clock calibration, control of software response variability, etc. In this sense, the TB task suits remote acquisition because of its low hardware requirements: it presents stimuli much longer than the time precision of computers. This task also allows the participants to respond without time pressure, differently from a reaction time task, for example, where the response time precision is critical.

Recently, Cravo and collaborators (2022) used the NeurUX platform (neurux.com) to acquire data online in a temporal production task. Here, we used the *OpenSesame* software in conjunction with the JATOS platform to implement the TB task remotely. These tools are relatively user-friendly and allow experimenters to program many psychophysical experiments. Furthermore, to the best of our knowledge, this is the first description of remote data acquisition using the temporal bisection task.

However, it is important to point out that the experimental conditions must be carefully observed, and the instructions to the participants must be carried out to minimize the most varied distractions and differences between each of the task execution locations. On the other hand, temporal perception occurs continuously during our daily lives, with the most varied distractions and differences in conditions (light, sound, humidity, etc.). Thus, the ecological validity of these studies may benefit from further scrutiny.

Because the remote and the live collection group share the same ratio TL/TS= 4, it is possible to compare the BP normalized by the AM and GM (Kopec et al., 2010). Similarly, the WR is comparable in experiments with different reference intervals because it is a normalized measure. Our remote data showed no significant differences from results obtained in live experiments regarding these comparable variables: BP/AM, BP/GM, and WR. The BP and DT were significantly different in the remote and live groups. We expected such a difference since the reference intervals were different, and hence the measures are not suitable for a direct comparison. In sum, we could not find any evidence that the modality of data acquisition (remote or live) influenced the results. Hence, the methodology described here may be a promising direction for increasing sample size, sample diversity, and reliability of conclusions.

5. ACKNOWLEDGMENTS

The authors of this study would like to express their gratitude to all colleagues from the research group, GEANM (Study Group on Neuropsychiatric Aspects and Motor skills), for their patient guidance, enthusiastic encouragement, and useful critiques of this research work. Also, we would like to thank the Federal University of ABC for providing a fellowship to MVA. We thank Peter Claessens for useful methodological discussions. We also thank Armando Machado for feedback on early versions of the manuscript.

6. Open Practices Statement

The data and materials for all experiments are available at <https://github.com/m-v-arruda/Temporal-Bisection-Task>, at the Supplementary Materials section and in the APPENDIX I. And none of the experiments were pre registered.

7. REFERENCES:

Andrade C. (2020). Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis: As Simple as It Gets. The Journal of clinical psychiatry, 81(5), 20f13681. https://doi.org/10.4088/JCP.20f13681

Allan, L & Gibbon, J. (1991). Human bisection at the geometric mean. Learning and Motivation. 22. 39-58. 10.1016/0023-9690(91)90016-2.

Block RA & Gruber RP. Time perception, attention, and memory: a selective review. Acta psychologica 2014; 149, 129-133.

Borsa, Juliane Callegaro, Damásio, Bruno Figueiredo and Bandeira, Denise Ruschel. Cross-cultural adaptation and validation of psychological instruments: some considerations. Paidéia (Ribeirão Preto) [online]. 2012, v. 22, n. 53, pp. 423-432.

Brucki, Sonia M.D. et al. Suggestions for the use of the mini mental state examination in Brazil.

Archives of Neuro-Psychiatry [online]. 2003, v. 61, n. 3B, pp. 777-781.

Campos, M. R., Schramm, J. M. D. A., Emmerick, I. C. M., Rodrigues, J. M., Avelar, F. G. D., & Pimentel, T. G. (2020). Burden of disease from COVID-19 and its acute and chronic complications: reflections on measurement (DALY) and perspectives in the Unified Health System. *Cadernos de Saúde Pública*, *36*. (11), 1-14.

Castro, ACV, Andréia, MPC, & Machado, KCA. The perception of time: contributions from the bisection procedure. Issues in Psychology 2013, 21(1), 49-70.

Catania, A. C. (1970). Reinforcement schedules and psychophysical judgment: A study of some temporal properties of behavior. The theory of reinforcement schedules.

Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. Journal of experimental psychology. Animal behavior processes, 3(3), 216–228. https://doi.org/10.1037//0097-7403.3.3.216

Cohen, J. (1988). Statistical power analysis for the behavioral sciences.

Cravo, A. M., Brito De Azevedo, G., Moraes, C., Azarias, B., Catheryne Barne, L., Dantas Bueno, F., de Camargo, R. Y., Carneiro Morita, V., Ventura, E., Sirius, P., Recio, R. S., Silvestrin, M., Machado De, R., Neto, A. (2022). Time experience during social distancing: A longitudinal study during the first months of COVID-19 pandemic in Brazil. In Sci. Adv (Vol. 8).

Edwards M. (2004). Assessing for depression and mood disturbance in later life. British journal of community nursing, 9(11), 492-494.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research, 12(3), 189-198.

Fotuhi M, et al. Neurobiology of COVID-19. Journal of Alzheimer's disease: JAD 2020; 76(1), 3-19.

Gallo Marin, B., Aghagoli, G., Lavine, K., Yang, L., Siff, E. J., Chiang, S. S., ... & Michelow, I. C. (2021). Predictors of COVID-19 severity: A literature review. Reviews in medical virology, *31*(1), 1-10.

Gil, S., & Droit-Volet, S. (2009). Time perception, depression and sadness. *Behavioural Processes*, *80*(2), 169–176. https://doi.org/10.1016/j.beproc.2008.11.012

Hinton, Danny & Stevens-Gill, Debbie. (2016). Online Psychometric Assessment. 10.1057/9781137517036_14. Applied Cyberpsychology: Practical Applications of Cyberpsychological Theory and Research.

International Test Commission. (2010). International Test Commission guidelines for translating and adapting tests. http://www.intestcom.org/upload/sitefiles/40.pdf

Jozefowiez, J., Gaudichon, C., Mekkass, F., & Machado, A. (2018). Log versus linear timing in human temporal bisection: A signal detection theory study. *Journal of Experimental Psychology: Animal Learning and Cognition*, *44*(4), 396–408. https://doi.org/10.1037/xan0000184

Jusyte, A., Schneidt, A., & Schönenberg, M. (2015). Temporal estimation of threatening stimuli in social anxiety disorder: Investigation of the effects of state anxiety and fearfulness. *Journal of Behavior Therapy and Experimental Psychiatry*, *47*, 25–33. https://doi.org/10.1016/j.jbtep.2014.11.006.

Katsuki F & Constantinidis C. Bottom-up and top-down attention: different processes and overlapping neural systems. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry 2014; 20(5), 509-521.

Kopec CD & Brody CD. Human performance on the temporal bisection task. Brain and cognition 2010; 74(3), 262-272.

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses.

Social Psychological and Personality Science. link, 1, 1-8.

Lange K, Kühn S, Filevich E (2015) Correction: “Just Another Tool for Online Studies” (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. PLoS ONE 10(7): e0134073.

Levy JM, Namboodiri VM & Hussain Shuler MG. Memory bias in the temporal bisection point. Frontiers in integrative neuroscience 2015; 9, 44.

Mathot, S., & March, J. (2021). Conducting linguistic experiments online with OpenSesame and OSWeb. PsyArXiv. February 10. doi:10.31234/osf.io/wnryc.

Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: An open-source, graphical experiment builder for the social sciences. Behav Res 44, 314–324 (2012).

Matthews WJ & Meck WH. Temporal cognition: Connecting subjective time to perception, attention, and memory. Psychological bulletin 2016; 142(8), 865-907.

Paniz-Mondolfi, A., Bryce, C., Grimes, Z., Gordon, R. E., Reidy, J., Lednicky, J., ... & Fowkes, M. (2020). Central nervous system involvement by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). Journal of medical virology, 92(7), 699-702.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. Journal of experimental psychology. Human perception and performance, 26(6), 1770-1787.

Roman, M. W., & Callen, B. L. (2008). Screening instruments for older adult depressive disorders: updating the evidence-based toolbox. Issues in mental health nursing, 29(9), 924-941.

Rosenthal, J. (1996). Qualitative Descriptors of Strength of Association and Effect Size. Journal of Social Service Research. 21. 37-59.

Stone, D. L., Deadrick, D. L., Lukaszewski, K. M., & Johnson, R. (2015). The influence of technology on the future of human resource management. Human Resource Management Review, 25(2), 216-231.

Stubbs, A. (1968). THE DISCRIMINATION OF STIMULUS DURATION BY PIGEONS. Journal of the Experimental Analysis of Behavior, 11(3), 223–238. https://doi.org/10.1901/jeab.1968.11-223

Terao, Y., Honma, M., Asahara, Y., Tokushige, S., Furubayashi, T., Miyazaki, T., Inomata-Terada, S., Uchibori, A., Miyagawa, S., Ichikawa, Y., Chiba, A., Ugawa, Y., & Suzuki, M. (2021). Time Distortion in Parkinsonism. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.648814

Thoenes, S., & Oberfeld, D. (2017). Meta-analysis of time perception and temporal processing in schizophrenia: Differential effects on precision and accuracy. Clinical psychology review, 54, 44–64. https://doi.org/10.1016/j.cpr.2017.03.007

The jamovi project (2021). jamovi. (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.

TIBCO Statistica (2020). TIBCO Software Inc, Palo Alto, CA, USA. (Version 14.0.0) [Computer Software]. Retrieved from: https://www.tibco.com/products/tibco-statistica

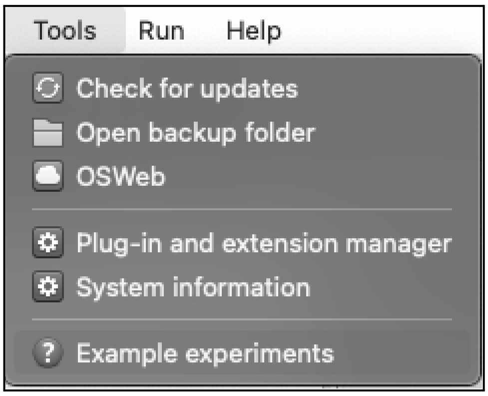
van Rijn, H., Gu, B. M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. Advances in experimental medicine and biology, 829, 75–99. https://doi.org/10.1007/978-1-4939-1782-2_5

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software. link, 36, 1-48.

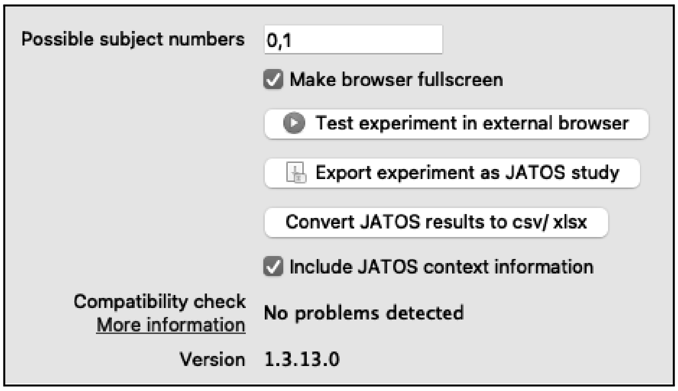
Wearden, J. (1991). Do Humans Possess an Internal Clock with Scalar Timing Properties?. Learning and Motivation. 22. 59-83. 10.1016/0023-9690(91)90017-3.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. Journal of psychiatric research, 17(1), 37–49. https://doi.org/10.1016/0022-3956(82)90033-4

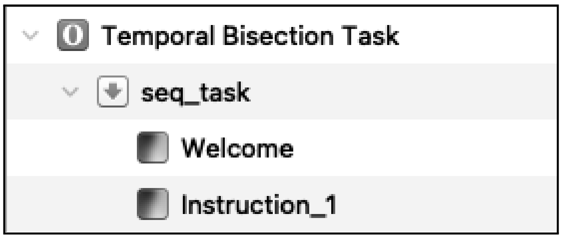
6. SUPPLEMENTARY MATERIALS:



Supplementary Figure 1 - Accessing the *OSWeb* tool.



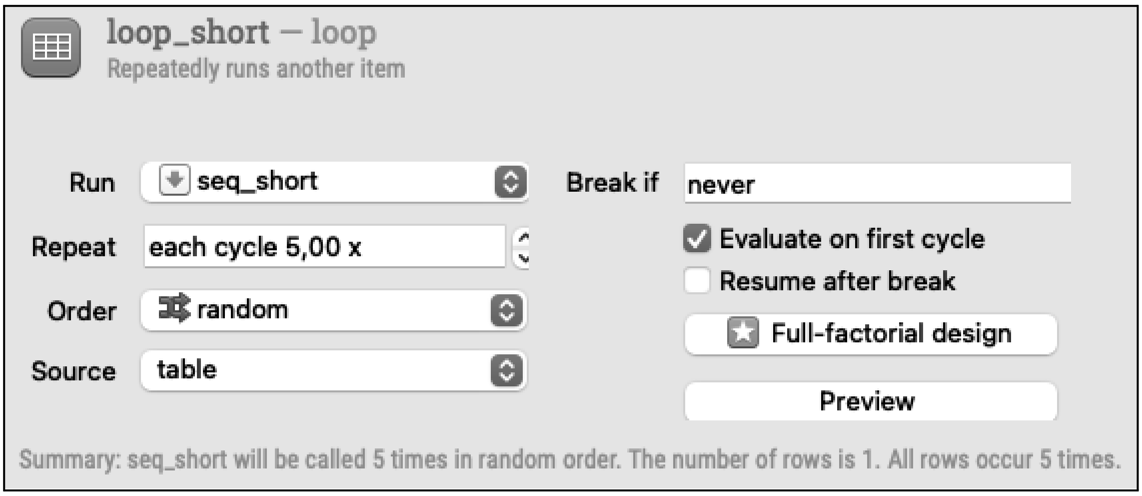
Supplementary Figure 2 - *OSWeb* tool page.



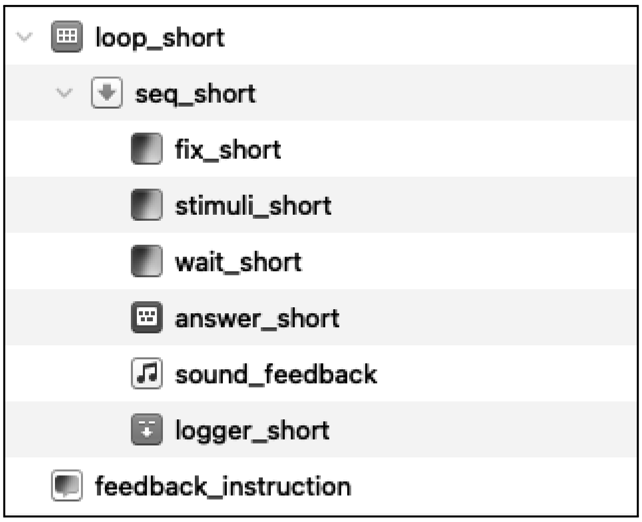
Supplementary Figure 3 - Starting an experiment in *OpenSesame.*

duration	congruency	answer_correct
500	short	q
2000	long	p

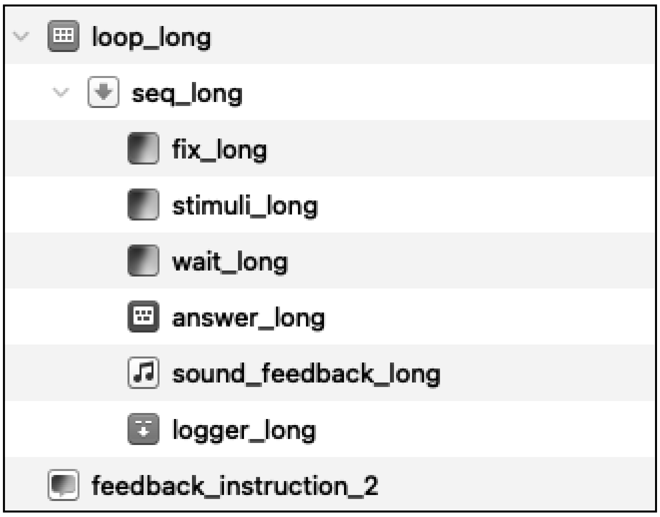
Supplementary Figure 4 - Example of a condition table from the 2nd Step.



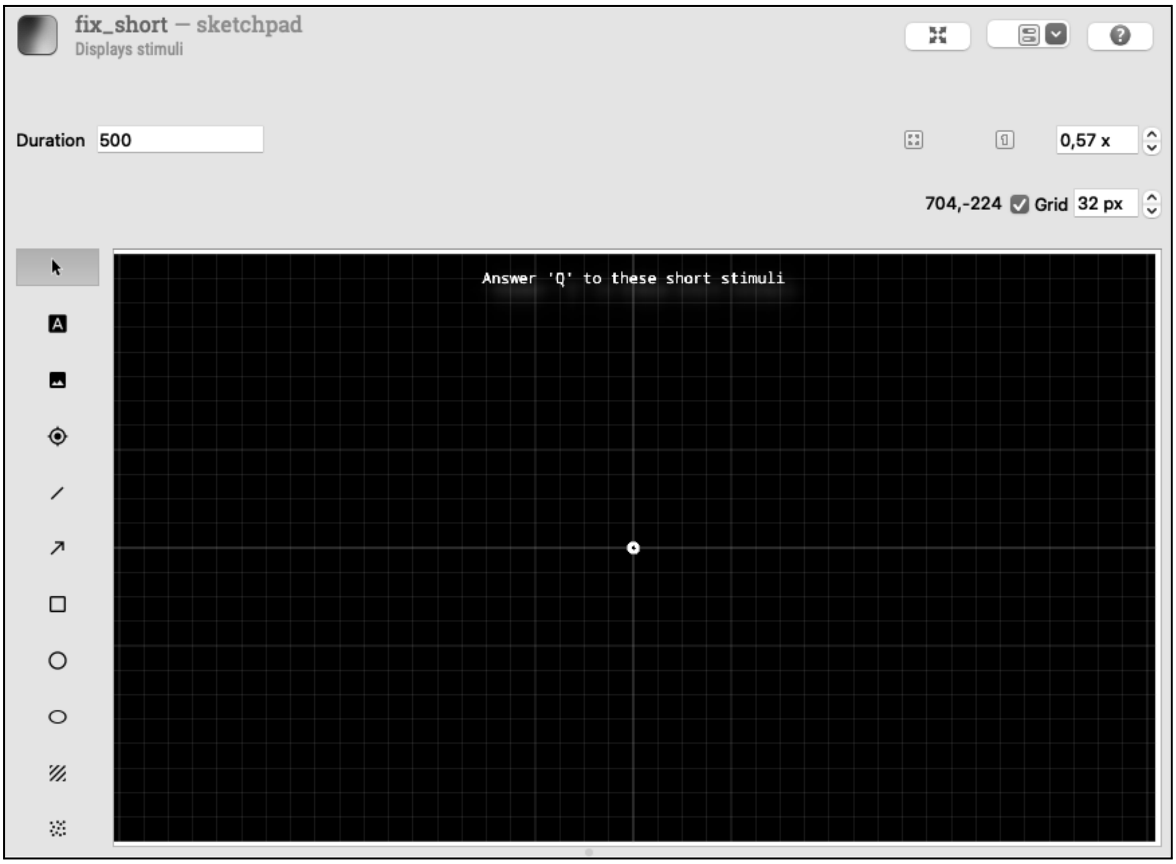
Supplementary Figure 5 - Example of the *loop* page.



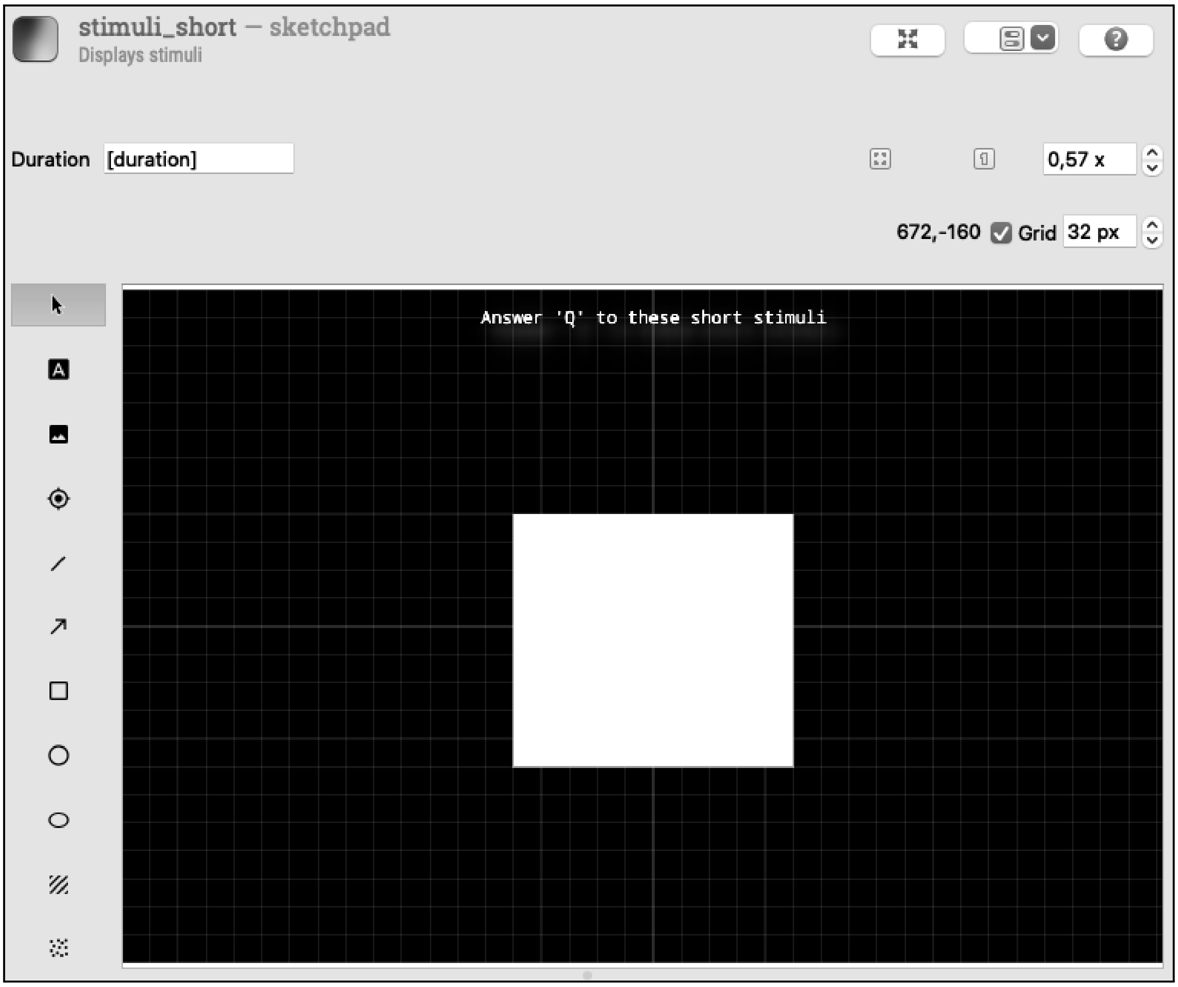
Supplementary Figure 6 - First part of 1st Step - short interval (500ms) stimulus.



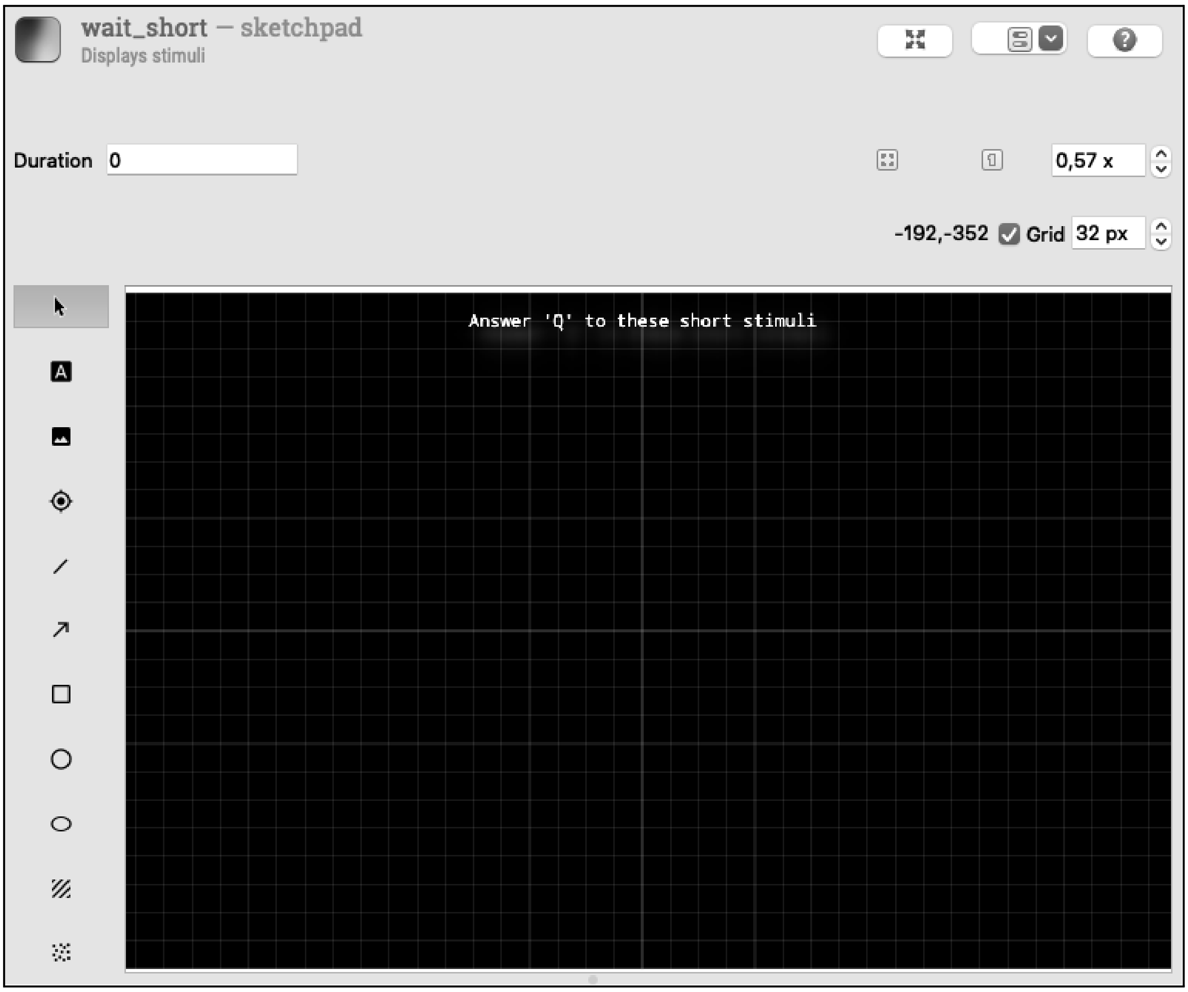
Supplementary Figure 7 - Second part of 1st Step - long interval (2000ms) stimulus.



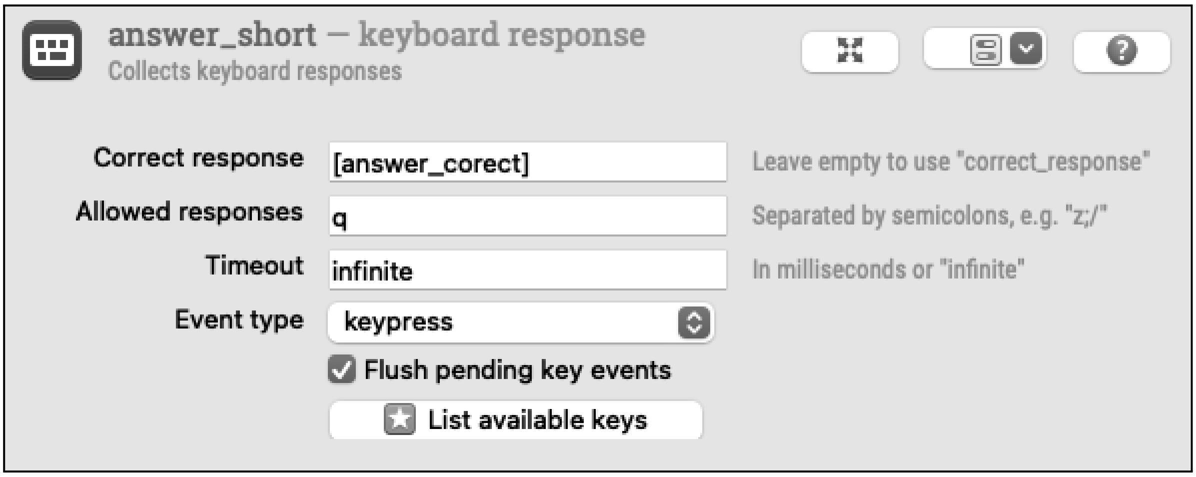
Supplementary Figure 8 - Example of the fixation page, with a duration of 500ms.



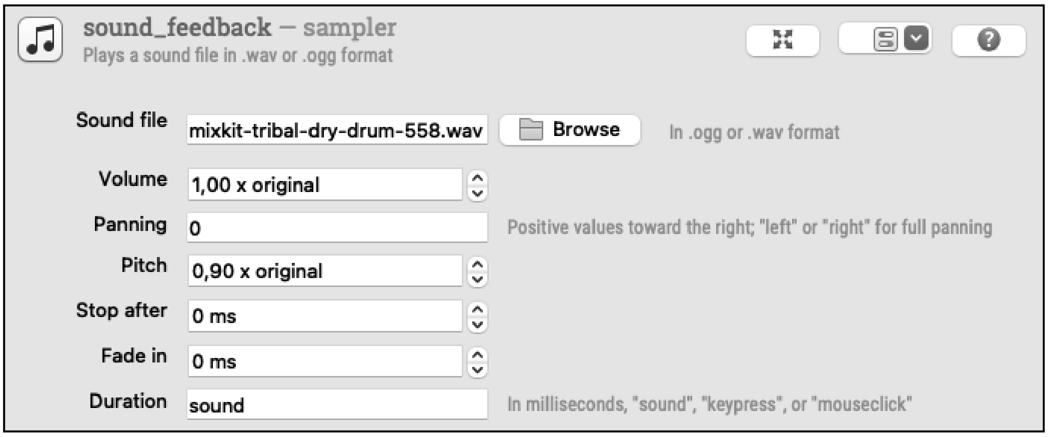
Supplementary Figure 9 - Example of the stimulus page, with duration corresponding to the table of conditions.



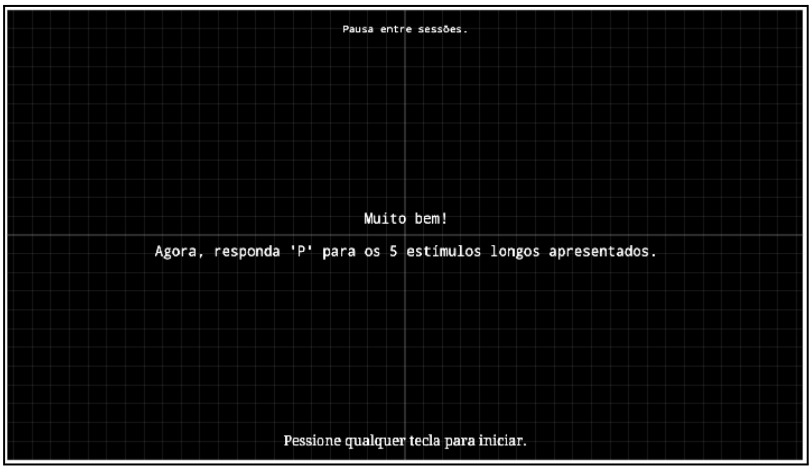
Supplementary Figure 10 - Example of the waiting page, with duration 0 (zero).



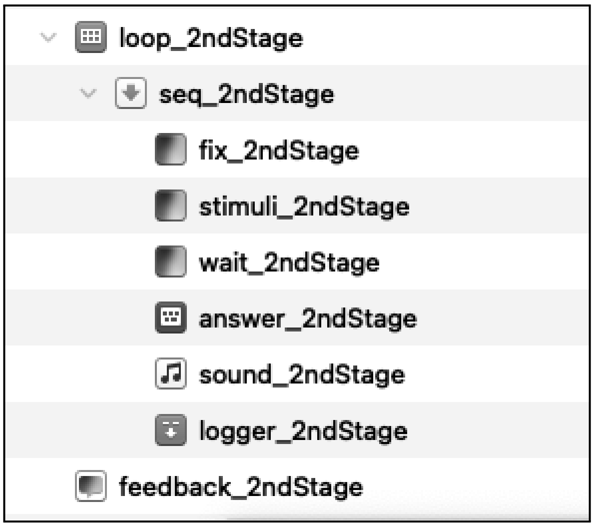
Supplementary Figure 11 - *Keyboard_response* page.



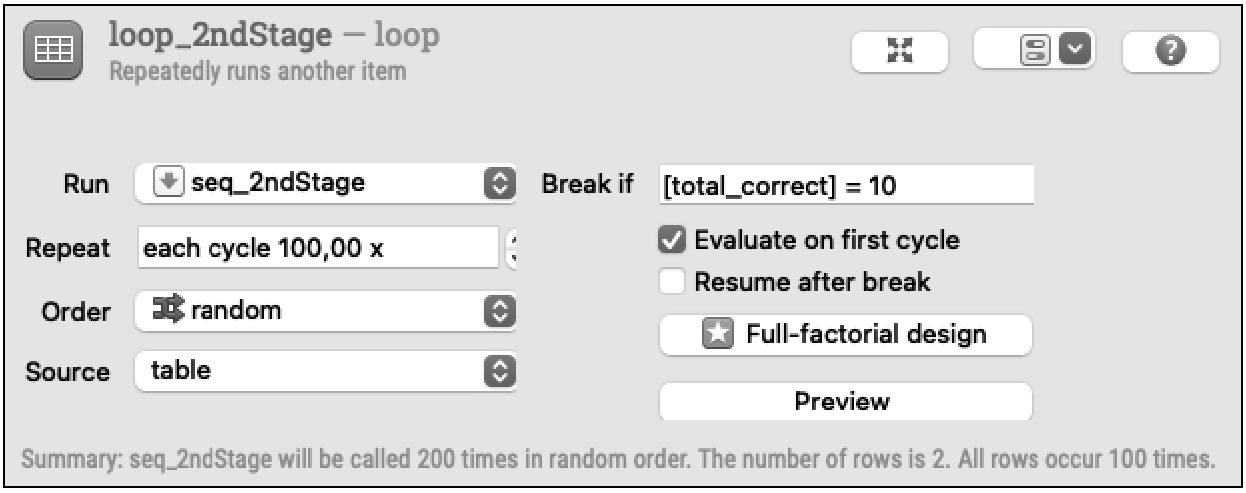
Supplementary Figure 12 - *Sampler* page.



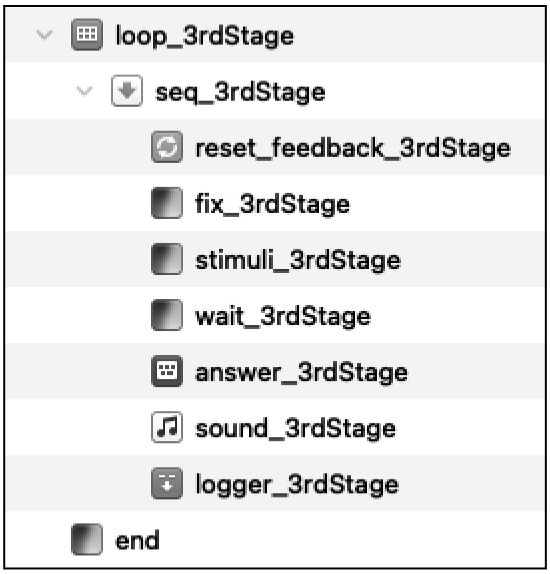
Supplementary Figure 13 - *Feedback* page.



Supplementary Figure 14 - Second phase: presenting the reference stimuli randomly.



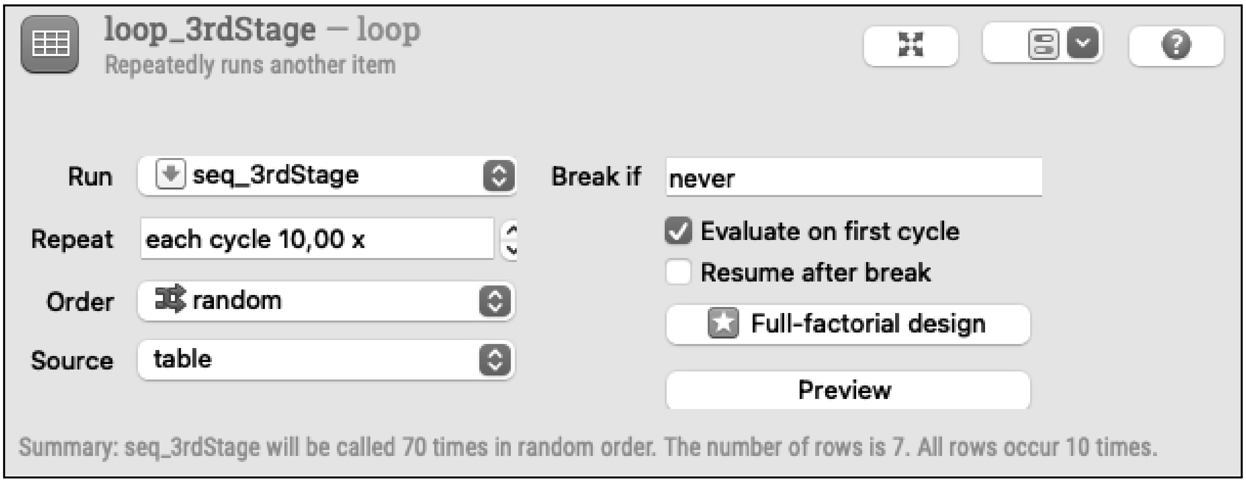
Supplementary Figure 15 - *Loop* properties of the second phase.



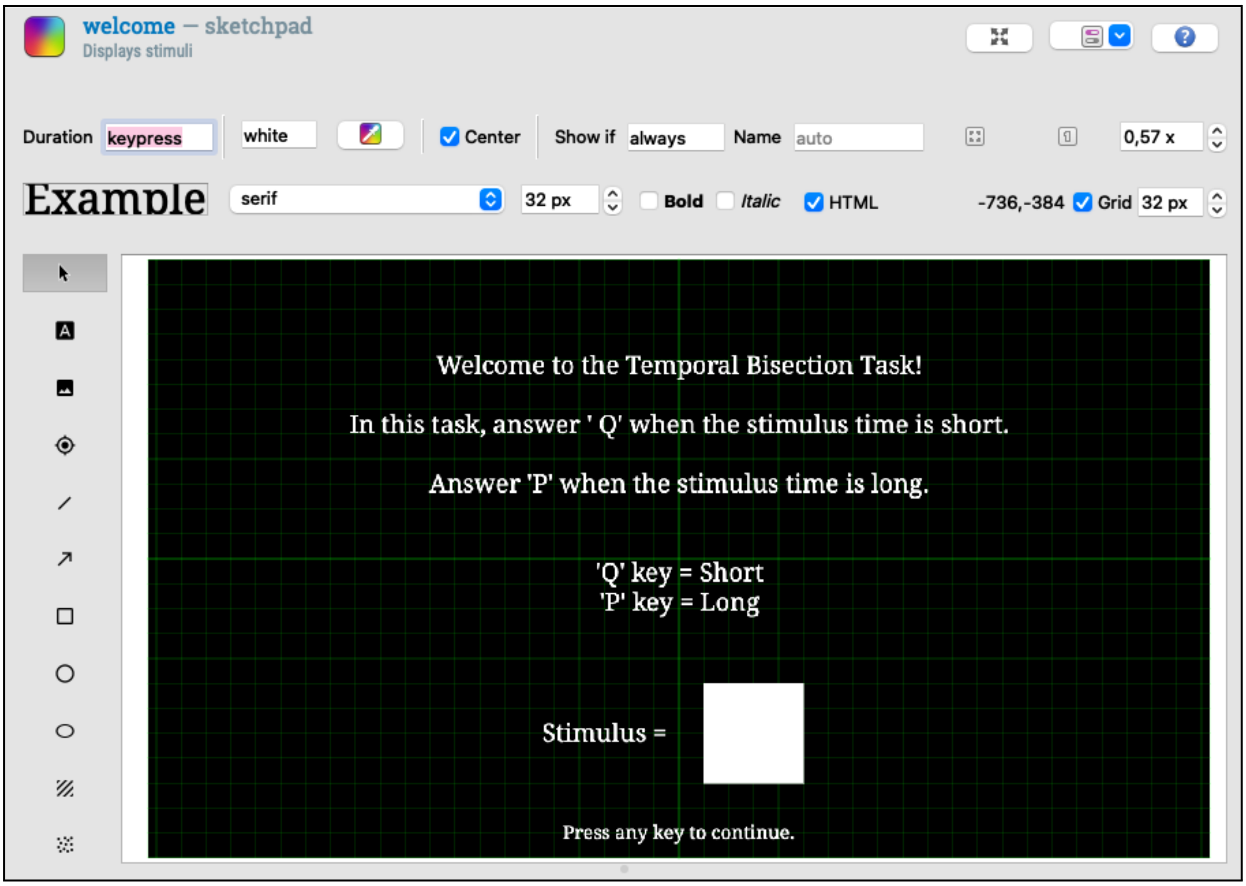
Supplementary Figure 16 - Third phase: Classification of intermediate stimuli.

duration_3rdStage
500
630
793.5
1000
1260
1587
2000

Supplementary Figure 17 - Condition table from 3rd Step.



Supplementary Figure 18 - Third phase *loop* properties.



Supplementary Figure 19 - Welcome page with initial instructions.

APPENDIX I - Algorithm

The psychometric function, its graphs, the BP, DT, and WR were obtained with custom-made *python 3.8* routines*,* using the *Google Colab* platform. The data obtained by the JATOS platform were converted to *.csv* format using the *OSWeb* tool of the *OpenSesame* software. Then the *.csv* spreadsheet containing the participants 'short' and 'long' responses was imported into *Google Colab* using the *pandas* library. We also used the *NumPy, matplotlib.pyplot*, and *scipy.optimize* libraries.

We computed the probability of responding long $P_\ell(T)$ using only the 70 test trials from Phase 3. For each interval, the probability was computed as the ratio between the number of long responses and the total number of responses. This is automatically computed with the *crosstab* routine from *pandas*. The fraction of 'short' or 'long' responses for each of the stimulus durations is stored in the *perc_answer* variable.

We fitted a sigmoid function to the presented stimuli (*time_stm*) as a function of P_ℓ (*perc_long).* The sigmoid curve was defined as:

$$f(x) = \frac{1}{(1 + e^{-a*(x-BP)})} . \qquad (i)$$

We used the *curve_fit* routine from *scipy.optimize* package to fit the curve, obtaining the optimized parameters (BP, DT and WR). The BP is equal to parameter *b*.

We obtained the differential threshold by calculating the intervals $T1_{(P\ell=0,25)}$ and $T2_{(P\ell=0,75)}$ directly from *a* and *b*:

$$T1 = b - \frac{log(3)}{a} \qquad (ii)$$

$$T2 = b + \frac{log(3)}{a} \qquad (iii)$$

The Weber ratio was calculated as (T2 - T1) / BP. The code used in the analysis can be accessed at the following link: <https://github.com/m-v-arruda/Temporal-Bisection-Task>.