The respiratory occlusion discrimination task: a new paradigm to measure respiratory interoceptive accuracy

Maaike Van Den Houte[1,4,6], Elke Vlemincx[2], Mathijs Franssen[3], Ilse Van Diest[3], Lukas Van Oudenhove[4,5] and Olivier Luminet[1, 6]

[1]Research group on Health, Emotion, Cognition, and Memory, Research Institute for Psychological Sciences, UCLouvain.

[2]Department of Health Sciences, Vrije Universiteit Amsterdam.

[3]Health Psychology Research Group, KULeuven.

[4]Laboratory for Brain-Gut Axis Studies (LaBGAS), KULeuven.

[5]Cognitive & Affective Neuroscience Lab, Dartmouth College.

[6]Fund for Scientific Research (FRS-FNRS).

Running head: the respiratory occlusion discrimination task

Corresponding author information:

Maaike Van Den Houte

UZ Herestraat 49 – box 7003, 3000 Leuven, Belgium

0032 16 37 46 95

maaike.vandenhoute@kuleuven.be

Abstract

Objective: Interoception, or the sense of the internal state of the body, is hypothesized to be essential for a wide range of psychobiological processes and the development and perpetuation of several (mental) health problems. However, the study of interoceptive accuracy, the objectively measured capacity to detect or discriminate conscious bodily signals, has been hampered by the use of tasks with questionable construct validity and is often limited to studying interoception solely in the cardiac domain.

Methods: We developed a novel task to measure interoceptive accuracy in the respiratory domain, the respiratory occlusion discrimination (ROD) task. In this task, interoceptive accuracy is defined as an individual's ability to detect small differences in lengths of short respiratory occlusions, assessed by means of an adaptive staircase procedure. This paper describes a validation study (N = 97) aimed at investigating the internal consistency, test-retest reliability and discriminant validity of the ROD task.

Results: The average just noticeable difference of lengths of respiratory occlusion was 74.22, with large inter-individual variability (SD = 37.1). The results of the validation study indicate acceptable internal consistency (Cronbach's alpha = 0.70), one-week test-retest reliability (r = 0.53) and discriminant validity, as indicated by a lack of correlation between the ROD task and an auditory discrimination task with identical design (r = 0.18), and a weak correlation with breathing behavior (r= -0.27).

Conclusions: The ROD task is a promising novel paradigm to study interoceptive accuracy and its role in various psychobiological processes and disorders.

The respiratory occlusion discrimination task: a new paradigm to measure respiratory interoceptive accuracy

1. Introduction

Interoception can be defined as "the overall process of how the nervous system senses, integrates, stores, and represents information about the state of the inner body" (Khalsa et al., 2018, p. 501). Interoception is thought to be crucial for a wide range of psychobiological processes, such as emotion recognition and regulation, symptom perception, and regulation of food intake (Craig, 2002). Interoceptive dysfunction is accordingly thought to be implicated in several psychiatric and somatic disorders (Khalsa et al., 2018) such as anxiety and mood disorders (Domschke, Stevens, Pfleiderer, & Gerlach, 2010; Harshaw, 2015; Van Diest, 2019), different types of functional somatic syndromes (Van den Bergh, Witthöft, Petersen, & Brown, 2017), obesity (Herbert & Pollatos, 2014), anorexia and bulimia nervosa (Jenkinson, Taylor, & Laws, 2018), and autism spectrum disorders (DuBois, Ameis, Lai, Casanova, & Desarkar, 2016). It is therefore not surprising that the study of interoceptive abilities and interoceptive dysfunction has boomed in recent years throughout different fields in (health) psychology, psychiatry and medicine (Khalsa & Lapidus, 2016).

Interoception can be conceptualized and measured on different levels inside and outside of conscious awareness. On a conscious level, Garfinkel, Seth, Barrett, Suzuki, & Critchley (2015) propose three dimensions of interoceptive abilities. Interoceptive accuracy refers to the objectively measured capacity to detect or discriminate bodily signals and is quantified by the performance on a behavioral task. Interoceptive sensibility refers to the self-reported capacity to detect or discriminate bodily signals, typically measured with a questionnaire or with confidence ratings. Interoceptive awareness refers to the meta-cognitive insight in interoceptive accuracy and is measured by relating the interoceptive accuracy score to confidence ratings. This paper involves the measurement of interoceptive accuracy.

Interoceptive accuracy is most often measured by comparing the perception of a physiological signal with an objective measure of that (natural or induced) physiological signal. The most often used task to measure interoceptive accuracy is the heartbeat counting task (HCT). This task requires individuals to count their heartbeats at rest during different time intervals, while heart rate is derived objectively from an electrocardiogram (Schandry, 1981), or with a pulse oximeter or smartwatch. Some major advantages of this task are that it is simple, short and requires little equipment. However, the measurement of cardiac interoceptive accuracy does not easily lend itself to signal detection or signal differentiation techniques (which are the golden standard in the long-standing tradition of exteroception research), because it is difficult to precisely manipulate cardiac signals in a non-invasive manner. This is problematic because it implies that the score is heavily influenced by pre-existing physiological differences between participants that are unrelated to interoception, such as the resting heart rate (Zamariola et al., 2018) and blood pressure (O'Brien, Reid, & Jones, 1998). Moreover, although interoception is often assumed to be a trait-like bodily domain-transcending ability, there is little evidence for this as performance measures of accuracy tasks in different bodily domains seem to be uncorrelated (e.g. Ferentzi et al., 2018; Garfinkel et al., 2016). Both issues highlight the need for the development of interoceptive tasks outside of the cardiac domain. Measuring interoceptive accuracy in the respiratory domain is therefore a valuable addition to the investigation of interoception. In the respiratory domain, it is more feasible to precisely manipulate the interoceptive signal experimentally and present experimenter-controlled interoceptive stimuli that are noticeable to all participants (compared to heartbeats at rest; Khalsa, Rudrauf, Sandesara, Olshansky, & Tranel, 2009).

Respiratory interoceptive accuracy is often measured using respiratory load detection or differentiation paradigms by adding respiratory resistive loads of different magnitudes to inspiration or expiration, making breathing more difficult (e.g. Webster & Colrain, 2000; Daubenmier, Sze, Kerr, Kemeny, & Mehling, 2013; Garfinkel et al., 2016; Schroijen, Davenport, Van den Bergh, &

Van Diest, 2020). Specifically, participants are asked whether an in-or exhalation contained a resistive load (detection paradigms), or are asked to judge which of multiple loaded breaths contained the strongest respiratory load (differentiation paradigms). It is easy to use a signal detection approach with these paradigms, and because loaded breathing resembles respiratory sensations encountered in daily life, such as upper respiratory congestion or face mask breathing, these tasks have high ecological validity. However, there are some downsides to using respiratory resistive loads as stimuli in an interoceptive accuracy task. First, when studying interoception one would prefer to use only neutral stimuli that do not have an intrinsic negative value (Van den Bergh, Bogaerts, & Van Diest, 2015). There is high inter-individual variability in the unpleasantness ratings of resistive loads (Von Leupoldt & Dahme, 2005), thus it is difficult to find a range of stimuli that are both noticeable and neutral for all subjects. Second, since only one respiratory load can be administered per breath, the trial duration in differentiation paradigms is very long. Both these pitfalls complicate the use of adaptive staircase paradigms, which are often used in exteroception research. A recently revised respiratory load differentiation task, the Filter Detection Task (Harrison et al., 2020, based on Garfinkel et al. (2016) and Harver et al. (1993)), bypasses these pitfalls by presenting only very low-level resistive loads using respiratory filters, and by using an adaptive task-performance algorithm allowing to reliably search for accuracy levels with a minimal number of trials. However, the error margin of respiratory resistances achieved by these filters may be larger than the differences in individuals' accuracy levels. A high precision of stimuli is necessary to introduce small enough differences and detection thresholds to study interoceptive accuracy and its correlates.

In this paper, we propose a new measure of respiratory interoceptive accuracy, the respiratory occlusion discrimination task (ROD task). In order to overcome the pitfalls of using resistive loads listed above, very short *inspiratory occlusions* (i.e. interruptions in inspiration) are used. More specifically, respiratory interoceptive accuracy is, in the proposed task, defined as individuals' capacity to discriminate the lengths of two inspiratory occlusions. Because participants

are asked to judge length, as compared to intensity of the stimuli, performance is likely less dependent on breathing behavior. Further, we expect that short (less than 1 second) respiratory occlusions will not be perceived as unpleasant. Finally, with the right equipment, the length of the occlusions can be very precisely controlled by the experimenter. Following the exteroception literature (Leek, 2001), an adaptive staircase paradigm was used to efficiently measure individuals' capacity to discriminate the lengths of two inspiratory occlusions. In the presented study, we investigated the internal consistency, discriminant validity and test-retest reliability of the ROD task.

2. Method

The primary goal of this study was to explore the internal consistency, discriminant validity and test-retest reliability of the ROD task. Discriminant validity was assessed by investigating the relationship between ROD task performance and performance on an auditory discrimination task. As a secondary goal, we explored the relationship between ROD task performance and trait positive and negative affect, interoceptive sensibility, dyspnea catastrophizing and alexithymia as measured with self-report questionnaires. Further, we explored the relationship between ROD task performance and the most frequently used measure of interoceptive accuracy, the HCT. However, it is not straightforward whether a significant association between HCT and ROD task performance is to be expected (discussed further below). The final procedure of the ROD task was established after performing a pilot study (see Supplementary material).

2.1. Participants and missing data

Participants were recruited through an online experiment management system (Sona Systems, Estonia) at the University of Leuven. Exclusion criteria were any self-reported psychiatric or chronical medical disorders, taking antidepressants, anxiety medication, beta-blockers, analgesics or sleep medication, pregnancy, and having taken part in the pilot study. Ninety-seven healthy individuals (28 men) participated in the study (mean age = 23.33, SD = 4.74). Fifty-eight participants

were native Dutch speakers and received instructions and questionnaires in Dutch, 39 participants were fluent English speakers and received the instructions and questionnaires in English. Participants received 20 euros for participation. The study was approved by the Social and Societal Ethics Committee of the University of Leuven. All participants gave written informed consent before participating. Data were collected between July and December 2019.

All participants who finished the ROD task in the first test session were included in the analyses (N=97). Some participants were excluded from certain analyses: 1) due to problems with the measurement of respiratory physiology, maximal inspiratory airflow could not be calculated for three participants; 2) because of technical issues that were only resolved later in data collection, we failed to collect ECG data for 22 participants, therefore, the heartbeat counting score could not be calculated for these participants ; 3)Three participants did not return for the second session, so they did not perform the ROD task twice and they did not fill out the trait questionnaires; 4) The ROD task could not be administered to one additional person in the second session because of technical problems, but this participant did fill out the trait questionnaires.

2.2. Apparatuses and physiological recordings

During the ROD task, participants wore a nose clip and headphones designed to block sounds to cancel out possible noise coming from the occluder system (Peltor Optime I). In the participants room, participants were breathing through a microbacterial filter with an integrated mouthpiece (Microgard IIB; Vyaire Medical, IL, USA). The filter was connected to a Y-shaped non-rebreathing valve (2730 Series; Hans Rudolph Inc., Shawnee, USA) of which the inspiratory port was connected to the experimenter room through Hytrel tubing and an opening in the wall. Between the filter and Y-shaped non-rebreathing valve, a sampling line was connected to record absolute pressure differences. The part of the breathing set-up in the participant room was hidden from sight by a sheet. In the experimenter room, the tube was fitted unto a pneumotachograph measuring airflow (Model 4813; Hans Rudolph Inc., Shawnee, USA) and a three-way T-shaped inflatable balloon-type valve

(8250A Series; Hans Rudolph Inc., Shawnee, USA). One of the balloons in this valve was permanently inflated during the experiment, the other one was used to deliver the occlusions. The valve was connected to an automated pneumatic inflatable balloon-type valve controller (8230AF Series; Hans Rudolph Inc., Shawnee, USA), which could be triggered by the computer parallel port through a TTL signals. Pressure changes for airflow and absolute pressure were transported to a pneumotachograph amplifier 1 (Series 1110, Hans Rudolph Inc., Shawnee, USA). All analog signals were digitized at 1000Hz using a National Instruments 6221 PCI card and 2111 BNC connector block (National Instruments). Airflow was continuously measured and monitored by the experimenter on a computer monitor. A schematic overview of the set-up and the used equipment can be found in Figure 1. The relationship between task performance and breathing behavior was explored as well. Respiratory flow was visually inspected and processed offline breath by breath with Matlab R2019b (Mathworks Inc, MA, USA). Because of the suspicion that task performance might be influenced by breathing depth because of a higher resonance of the occlusions, the maximal airflow during inspiration was determined in every breath, both during the baseline measurement and during the respiratory occlusion discrimination task. Three measures were explored as possible confounders of task performance: average maximal inspiratory flow at rest, average maximal inspiratory flow during the breaths right before the breaths in which the occlusions were presented, and the average ratio of inspiratory flow right before the first versus the second occlusion during the respiratory occlusion discrimination task.

Heart rate was measured using an electrocardiogram (ECG) recording. Three disposable 8-mm Ag/AgCl ECG electrodes (Kendall H66LG, Medtronic, Ireland) were attached underneath the participants' clavicles and at the left lower ribs. The electrodes were positioned in a Lead II configuration. A Coulbourn V75-04 differential isolated amplifier with bandpass filter (Coulbourn Instruments; Allentown, PA, United States) was used to amplify and filter the signal (low-pass 150 Hz). The signal was digitized at 1000 Hz using the same data acquisition card as for the respiratory

signals. The ECG signal was visually inspected and processed offline with Matlab R2019b (Mathworks Inc, MA, USA).

2.3. Measures

2.3.1. Respiratory occlusion discrimination task (Figure 2): A transformed adaptive staircase paradigm (Leek, 2001) was used to investigate participants' capacity to detect differences in length between two occlusions. Adaptive staircase methods allow us to efficiently search for a specific point in the psychometric function, such as the just noticeable difference. A two-down, one-up procedure was used in order to approach the 70.7% correct differentiation point in the psychometric function (Levitt, 1971), meaning that we were looking for differences in occlusion lengths that would be noticeable to the participant about 70% of the time (defined further as the just noticeable difference (JND)). In a two-interval forced choice task, participants were presented with an occlusion pair on each trial, consisting of a reference occlusion (always 440ms long, including inflation and deflation time of the balloon through which the occlusion was presented) and a test occlusion, the latter being shorter or longer than the reference occlusion. The two occlusions were presented within one inspiration in random order, separated by an interval of 300ms. The researcher followed the participant's breathing pattern visually and manually triggered the start of the trial at the beginning of an inspiration. The inter-trial interval was approximately ten seconds – at least one occlusion-free breath was left in between two trials. The length of the test occlusion was dependent on the participant's answer on the previous trial: if the participant could successfully distinguish the test occlusion from the reference occlusion two times in a row, the length of the test occlusion became more similar to the length of the reference conclusion. However, if the participant once failed to distinguish the occlusions, the length of the test occlusion became less similar to the length of the reference occlusion.

An example of the staircase procedure can be found in Figure 2. The paradigm included both an upwards going staircase (approaching the reference occlusion with test occlusions shorter than the reference occlusion) and a downwards going staircase (approaching the reference occlusion with test occlusions longer than the reference occlusion). The staircases were presented intertwined, such that trial N belonged to the downwards staircase, and trial N+1 belonged to the upwards staircase. The maximum test occlusion length was 620ms, while the minimal occlusion length was 260ms. The step size, i.e. the duration by which the occlusion length is reduced/increased after two right/one wrong answers, varied throughout the experiment in order to first efficiently search for the participant's rough ability, and then fine-tune this. At the start of the experiment, the lengths were increased / decreased with 30ms. With every reversal, the step size was decreased with 5ms until a minimal step size of 5ms had been reached. A reversal can be defined as a change in direction of a staircase, e.g. after an incorrect answer following a series of correct answers (depicted by the arrows in Figure 2). The procedure ended when both (intertwined) staircases had reached six reversals. ROD task performance was defined as the Just Noticeable Difference (JND), which was calculated as the average of the reversal points. This procedure was established after subjecting a first version of the ROD task (based on the exteroception literature) to a pilot study, with the aim of investigating the feasibility, unpleasantness and efficiency of the procedure. The methods, results and conclusion of this pilot study can be found in the *Supplementary material.*

2.3.2. Auditory discrimination task: During the auditory discrimination task, a transformed adaptive staircase paradigm was used to investigate participants' capacity to detect differences in the lengths of neutral tones, which were presented through headphones. In a two-interval forced choice task, participants were on each trial presented with a pair

of neutral tones, consisting of a reference tone (always 440ms long) and a test tone (min 260ms long and max 620 ms long). Staircase rules and procedures were identical to those of the respiratory occlusion discrimination task. The tones were presented 300 ms apart and the inter-trial interval (time between participants' answer and the presentation of the next pair of tones) was 4500ms.

2.3.3. Heartbeat counting task (HCT): The HCT was used to measure cardiac interoceptive accuracy. Participants were instructed to sit still with their hands on their lap and to count, without feeling their pulse, their heartbeats in three intervals (25 sec, 35 sec., 45 sec) presented in random order. The start and end of the intervals was signaled with a neutral tone. After every interval, participants reported the counted number of heartbeats by typing them into the keyboard. To reduce the influence of beliefs about heartrate and time estimation, adapted instructions, as proposed by Desmedt, Luminet, & Corneille (2018) were used. Participants were explicitly asked to only count the heartbeats they actually felt, and not to estimate how many heartbeats they felt. Heartbeat counting scores were calculated using the standard formula: $1/3\Sigma(1-(|\text{actual heartbeats} - \text{reported heartbeats}|)/\text{actual heartbeats})$.

2.3.4. Trait Questionnaires: *Positive and negative affectivity* was measured with the trait version of the Positive And Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988; Dutch validation by Engelen, De Peuter, Victoir, Van Diest & Van den Bergh, 2006). Participants indicate how often they experience 10 negative and 10 positive emotions in daily life on a scale from 1 (not at all) to 5 (very much). *Dyspnea catastrophizing* was measured with the Breathlessness Catastrophizing Scale (BCS; Solomon et al., 2015). Participants indicate to what extent they experience 13 catastrophizing thoughts when they are experiencing breathlessness on a scale from 0 (not at all) to 4 (all the time). *Habitual symptom reporting* was measured with the

Checklist for Symptoms in Daily Life (CSD; Walentynowicz, Witthöft, Raes, Van Diest, & Van den Bergh, 2018). Participants indicate how often they have experienced 39 symptoms in the past year on a 5-point scale (never – very often). *Interoceptive sensibility* was measured with the Interoceptive Sensibility and Attention Questionnaire (ISAQ; Bogaerts et al., *manuscript in preparation*). Respondents indicate on a 5-point Likert scale to what extent each of 19 statements applies to them (1: completely disagree – 5: completely agree). The scale measures two facets of interoception: "interoceptive sensibility" (self-reported awareness of non-aversive bodily processes) and" attention to unpleasant bodily sensations" (self-reported tendency to focus on aversive bodily processes). *Alexithymia* was measured with the Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994). The TAS-20 consists of 20 items that can be answered on a 5-point Liker scale (1: completely disagree – 5: completely agree), measuring three subscales: Difficulty Identifying Feelings (DIF), Difficulty Describing Feelings (DDF) and Externally-Oriented Thinking (EOT).

2.4. Procedure

Affect 5 software (e.g., Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans (2010)) was used for stimulus presentation, experiment control and recording of the analog input. Participants came to the lab for two separate sessions, 6-8 days apart. In the first session, participants performed the HCT, ROD task and the auditory discrimination task. In the second session, they repeated the ROD task and filled out the trait questionnaires. Session 1. After signing the informed consent form, electrodes were attached for the ECG measurement and participants performed the HCT. The ECG electrodes were removed after the HCT. Next, participants were instructed to breathe through the respiratory system for two minutes to get used to the sensation of the mouthpiece and the nose clip and for the baseline measurement of respiratory flow. After this, participants received the task instructions: they would always feel a pair of occlusions of unequal length and had to indicate, with

the keyboard, which one of the two was longest. Subsequently, participants completed a practice phase consisting of eight trials with test occlusion lengths of 290, 335, 350, 360, 520, 530, 545, and 590ms presented in random order. As some people have the reflex to cease their inspiration when they feel an occlusion, participants were explicitly instructed to not stop breathing but keep inhaling until at least the end of the second occlusion. During this practice phase, participants received visual feedback on whether their answer was correct for every trial. Next, participants completed the staircase procedure as described above. The staircase was interrupted after every 15 trials so that participants could take out the mouthpiece and take a short break (anywhere between 30 and 120 seconds at their own choice). After the ROD task ended (i.e. when six reversals had been reached for both the upwards and downwards going staircase), participants rated on a scale from 1-10 how difficult they found the task, how focused they were on the task, and how motivated they felt to perform the task well. After the ROD task, participants practiced the auditory discrimination task: eight trials with test tone lengths of 290, 335, 350, 360, 520, 530, 545, and 590 ms were presented in random order. During this practice phase, participants received visual feedback on whether their answer was correct for every trial. Finally, participants completed the staircase procedure for the auditory discrimination task. Session 1 lasted 60-75 min. Session 2: Participants signed the informed consent form and performed the ROD task a second time (including baseline measurement of respiratory flow and practice phase). Next, they filled out the trait questionnaires (PANAS, CSD, IAQ and TAS-20) through the online survey system Qualtrics. Session 2 lasted 30 – 45 minutes.

2.5. Data analysis

Data analysis was performed with SAS 9.4 (SAS Institute, Cary, NC, USA). Only the results of the ROD task of the first session were used to investigate task performance, internal consistency and discriminant validity. The internal consistency was investigated by calculating the Cronbach's alpha of the downwards JND (average difference between the six reversal points in the downwards staircase and the reference occlusion length) and the upwards JND (average difference between the

reference occlusion length and the six reversal points in the upwards staircase). ROD task performance was defined as the overall JND, i.e. the average between the downwards and upwards JND. A higher JND means a higher difficulty to distinguish the test occlusions from the reference occlusions and thus worse task performance and lower interoceptive accuracy. Discriminant validity was investigated by calculating the Pearson correlation between ROD task performance and task performance on the auditory discrimination task. Further, because it is conceivable that the features (such as the length) of the occlusions are processed more easily by participants who breathe in more deeply we explored whether ROD task performance was correlated with breathing behavior during baseline and during the task. This was done because the variance in ROD task performance should predominantly reflect inter-individual differences in interoceptive accuracy, and not individual differences in breathing behavior. Finally, test-retest reliability was investigated by calculating the correlation between the ROD task performances in sessions 1 and 2. As a secondary goal, we explored the correlation between the self-report trait questionnaires and ROD task performance as well as the correlation between HCT and ROD task performance. P-values of the correlations between ROD task performance and trait questionnaires were corrected for multiple testing with the False Discovery Rate method (FDR; Benjamini & Hochberg, 1995).

Results

3.1. Task performance

The average JND was 74.22 (SD = 37.1). This means that on average, participants could distinguish test occlusions differing 74.22ms from the reference occlusions about 70% of the time. The distribution of ROD task performance was positively skewed. Therefore, the scores were logarithmically transformed before further analysis. There was no difference in task performance between Dutch speakers and speakers who received the instructions in English ($t_{95} = 0.09$, $p = 0.93$) and between men and women ($t_{95} = 0.97$, $p = 0.34$). The average number of trials needed to end the

staircase procedure was 61.94 (SD = 13.4). A description of the average trial number and occlusion length at each reversal point for both the upwards and downwards going staircase can be found in Table 1. The time necessary to complete the staircase procedure depended on the number of trials it took to reach six reversals and the break length chosen by the participant, but was on average 14.69 minutes (SD = 4.15, range 4.53 – 26.19 minutes). ROD task performance was uncorrelated to self-reported motivation ($r = -0.16$, $p = 0.11$) and focus (-0.02, $p = 0.79$), but was correlated to perceived task difficulty, with participants who rated the task as more difficult having a higher JND and thus worse ROD task performance ($r = 0.22$, $p = 0.034$).

3.2. Internal consistency

It was easier for participants to differentiate the reference occlusion from occlusions that were shorter than the reference occlusion, compared to differentiating the reference occlusions from occlusions that were longer than the reference occlusion (average JND upwards going staircase = 68.13, SD = 42.25 average JND downwards going staircase = 80.31, SD = 42.44). This difference was significant ($t_{96} = 2.94$, $p = 0.004$). This was expected, since perceived differences in length will be larger at relatively shorter stimulus lengths (cfr. Weber's law). There was a medium-large correlation between the downwards JND and the upwards JND ($r = 0.53$, $p < 0.0001$). The Cronbach's alpha was 0.70, which is generally considered acceptable (Kline, 1999).

3.3. Discriminant validity

The average JND of the auditory discrimination task was 63.99 (SD = 23.11). This means that on average, participants could distinguish test tones differing 63.99ms from the reference tone about 70% of the time. There was a no significant correlation between performance on the auditory discrimination task and ROD task performance ($r = 0.18$, $p = 0.063$). Given that the only difference between the auditory discrimination task and the ROD task was the used stimuli, this suggests strong discriminant validity of the ROD task in the measurement of *interoceptive* discrimination.

3.4. Relationship with breathing behavior

ROD task performance was not significantly correlated to the average maximal inspiratory flow during rest ($r = -0.16$, $p = 0.12$), nor with the average maximal inspiratory flow during the breaths right before the breaths in which the occlusions were presented ($r = -0.10$, $p = 0.34$). ROD task performance was significantly correlated with the average ratio of inspiratory flow right before the first versus the second occlusion during the ROD task ($r = -0.27$, $p = 0.008$). This means that task performance was slightly better for those participants with higher inspiratory depth during the second occlusion relative to the first occlusion within one trial – or, alternatively, slightly worse for participants with reduced inspiratory depth during the second occlusion relative to the first one. Consequently, it is recommended to control for this latter variable when using the ROD task to assess respiratory interoceptive accuracy.

3.5. Test-retest reliability

The average task performance in the second session was 74.22 (SD = 37.1). This did not differ from the performance in the first session ($t_{92} = 0.90$, $p = 0.37$). There was a medium-large correlation between ROD task performance in sessions 1 and 2 ($r = 0.53$, $p < 0.0001$), suggesting good test-retest reliability.

3.6. Exploratory: correlation with HCT performance

The average heartbeat counting score was 0.31 (SD = 0.21). Performance on the HCT was not significantly correlated with ROD task performance ($r = -0.16$, $p = 0.18$).

3.7. Exploratory: correlations with trait questionnaires

Because of the significant correlation between task performance and breathing behavior and the marginally significant correlation between ROD task performance and auditory discrimination task performance, the exploratory correlations between ROD task performance and the trait

questionnaires were calculated (1) without controlling for any possible confounding variables, (2) while controlling for the average ratio of inspiratory flow right before the first versus the second occlusion during the respiratory occlusion discrimination task, (3) while controlling for performance on the auditory discrimination task, and (4) while controlling for both. Almost all of these correlations were close to zero, and none of them reached significance, even before correcting the p-values for multiple testing. There was a positive correlation between the JND of the ROD task and the Breathlessness Catastrophizing Scale (i.e. dyspnea catastrophizers performed worse on the ROD task), but this correlation only reached significance when controlling for performance on the auditory discrimination task and the significance did not survive FDR correction for multiple testing. Descriptive statistics of the questionnaire scores and detailed correlation results can be found in Table 3.

4. Discussion

The goal of this study was to validate a new task to measure interoceptive accuracy in the respiratory domain, the respiratory occlusion discrimination task (ROD task). In this task, interoceptive accuracy is defined as one's ability to reliably discriminate lengths of short respiratory occlusion. An adaptive staircase paradigm is used to efficiently find the test length at which the participant is able to successfully discriminate it from a reference length 70.7% of the time (Just Noticeable Difference; JND). Twenty-five healthy individuals participated in a pilot study (see supplementary material). The pilot study confirmed that the respiratory occlusions were perceived as neutral and that the procedure did not induce any anxiety in healthy individuals. Further, we established that the initially proposed procedure, which was based on the exteroception literature, was efficient in determining the JND but could be shortened. The final ROD task, which was slightly adapted based on the results of the pilot study, was then performed by 97 healthy individuals (who did not participate in the pilot study) in a validation study, aimed at investigating the internal consistency, discriminant validity and test-retest reliability.

The results demonstrated acceptable internal consistency and test-retest reliability (with the retest at one week) of the task. The most challenging feature of any task measuring interoceptive accuracy is however construct validity: does the task actually measure objective interoceptive abilities? Construct validity is conceivable if the tasks is highly correlated with other tasks measuring interoceptive accuracy (convergent validity) and uncorrelated to other measures that theoretically could explain the variance in task performance but are unrelated to interoceptive accuracy (discriminant validity). Convergent validity is difficult to establish for the ROD task, since there is currently no convincing benchmark respiratory accuracy task to compare it with (see introduction). The most often used task to measure interoceptive accuracy in the literature is the HCT. Although performance on the ROD task was correlated in the expected direction with performance on the HCT (note that a *small* JND notes better performance on the ROD task, while a *high* heartbeat counting score indicates better performance on the HCT), this relationship was not significant. However, this says little of the construct validity of the ROD task since 1) the construct validity of HCT itself is highly questionable, given its dependence on task instructions, time estimation capacities, knowledge about heart rate, resting heart rate and blood pressure (Desmedt et al., 2018; Meissner & Wittman, 2011; Brener & Ring, 2016, Zamariola et al., 2018; O'Brien et al., 1998) and 2) there is very little evidence that interoceptive accuracy is a trait-like domain-transcending ability (Ferentzi et al., 2018; Garfinkel et al., 2016). Even if the HCT, or any other interoceptive task outside the respiratory domain for that matter, would have high construct validity, there is no guarantee that it would correlate with an interoceptive task in the respiratory domain. If anything, the lack of correlation between performance on the ROD task and the HCT task demonstrates the urgent need for the development of valid tasks measuring interoceptive accuracy in various bodily domains, such as the cardiac domain and the gastrointestinal domain, in order to further elucidate the specificity vs. generalizability of interoceptive abilities across bodily domains and the relationship between

interoceptive abilities in different bodily domains and specific psychobiological functions and psychiatric disorders.

With regards to discriminant validity, the ROD task was compared to an auditory discrimination task using the exact same procedure with auditory stimuli instead of respiratory occlusions. An auditory stimulus was chosen because in the pilot study, 20% of participants indicated that they "listened" to the occlusions resonating in their mouth and throat in order to better be able to discriminate their lengths. However, when measuring interoceptive accuracy, it is important that the task does not measure exteroceptive accuracy instead. Given that performing the auditory discrimination task requires, except for interoception, similar resources (general attention and discrimination capacities, time estimation) as the respiratory occlusion discrimination task, a positive correlation was expected. The fact that this correlation was insignificant (with an explained variance of 3%) supports the idea that performance on the ROD task largely reflects interoceptive processes. As another possible confounder of task performance, we aimed to investigate whether breathing behavior is correlated with task performance. It is conceivable that the features (such as the length) of the occlusions are processed more easily by participants who breathe in more deeply, because the difference between the physiological states of occlusion and not-occlusion is more salient. There was indeed a small but significant positive correlation between breathing behavior during the task and task performance. Some participants had the tendency to reduce their breathing strength after the first occlusion within a trial, which might have made processing of the second occlusion harder. To reduce inter-individual variability in this behavior, we instructed participants before the start of the staircase procedure to not stop breathing but keep inspiring until at least the end of the second occlusion. However, we found that participants with a higher tendency to do this have worse task performance and thus lower interoceptive accuracy. Although this breathing behavior only explained 7% of the variation in ROD task performance, it is worth considering to control for this factor when using the

ROD task for investigating the relationship between interoceptive accuracy and other psychobiological functions.

The results of the pilot and validation studies suggest that the ROD task is a promising alternative to measure interoceptive accuracy in the respiratory domain. It has multiple advantages over existing interoceptive accuracy paradigms in the respiratory domain. First, the stimuli are precise and under complete control of the experimenter. This precision allows for a large inter-individual variability in the outcome measure (i.e. the JND), which is necessary when investigating the relationship between interoceptive accuracy and (mental) health processes. Second, the stimuli are non-aversive. This is important because in most definitions of interoception, the concept primarily refers to the sensory-perceptual component of the evaluation of bodily states, and not the affective-motivational component (Van den Bergh et al., 2015). Finally, the task has good internal consistency, test-retest reliability and discriminant validity.

However, the ROD task also suffers from some limitations that should be acknowledged and taken into account. First, the ROD task is more time-and labor-intensive than some other interoceptive accuracy tasks, such as the HCT. Although the ROD task can be largely automated, in its current form the experimenter manually triggers the pair of occlusions while inspecting breathing behavior because they need to be given at the right time during inspiration. Relatedly, in contrast with the HCT, advanced equipment to deliver the occlusions and to measure respiratory flow is required to administer the ROD task. Second, in our sample, the outcome score was not normally distributed with a slight positive skew. However, we feel that these limitations are largely outweighed by the advantages of the task. Finally, we should be aware that the ROD task is limited to measuring interoceptive accuracy in the respiratory domain and should not be viewed as a measure of "general interoceptive accuracy". It is still unknown whether interoceptive accuracy is a domain-transcending trait, although this is a highly relevant question. Therefore, researchers studying interoception should continue searching for validated tasks in different bodily domains.

In conclusion, the studies discussed here demonstrate the large potential of the ROD task as a new task to measure interoceptive accuracy in the respiratory domain, and opens up new possibilities in the widely applied study of interoception.

References

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23-32. https://doi.org/10.1016/0022-3999(94)90005-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 289-300.

Bogaerts, K., Walentynowizc, M., Van Den Houte, M., Constantinou, E., & Van den Bergh, O. (2019). *The Interoceptive Awareness Questionnaire (IAQ) differentiates between and within groups with stress-related bodily complaints versus healthy controls.* Paper in preparation.

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: the measurement of heartbeat detection. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160015. https://doi.org/10.1098/rstb.2016.0015

Craig., A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience, 3(8),* 655-666. https://doi.org/10.1038/nrn894

Daubenmier, J., Sze, J., Kerr, C. E., Kemeny, M. E., & Mehling, W. (2013). Follow your breath: respiratory interoceptive accuracy in experienced meditators. *Psychophysiology, 50*(8), 777-789. https://doi.org/10.1111/psyp.12057

Davenport, P. W., Chan, P. Y. S., Zhang, W., & Chou, Y. L. (2007). Detection threshold for inspiratory resistive loads and respiratory-related evoked potentials. *Journal of Applied Physiology*, *102*(1), 276-285. https://doi.org/10.1152/japplphysiol.01436.2005

Desmedt, O., Luminet, O., & Corneille, O. (2018). The heartbeat counting task largely involves non-interoceptive processes: Evidence from both the original and an adapted counting task. *Biological Psychology*, *138*, 185-188. https://doi.org/10.1016/j.biopsycho.2018.09.004

Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. *Clinical Psychology Review*, 30(1), 1-11. https://doi.org/10.1016/j.cpr.2009.08.008

DuBois, D., Ameis, S. H., Lai, M. C., Casanova, M. F., & Desarkar, P. (2016). Interoception in autism spectrum disorder: A review. *International Journal of Developmental Neuroscience*, 52, 104-111. https://doi.org/10.1016/j.ijdevneu.2016.05.001

Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van den Bergh, O. (2006). Verdere validering van de "Positive and Negative Affect Schedule" (PANAS) en vergelijking van twee Nederlandstalige versie. *Gedrag en Gezondheid*, 34, 89-102.

Ferentzi, E., Bogdány, T., Szabolcs, Z., Csala, B., Horváth, Á., & Köteles, F. (2018). Multichannel investigation of interoception: sensitivity is not a generalizable feature. *Frontiers in Human Neuroscience*, *12*, 223. https://doi.org/10.3389/fnhum.2018.00223

Garfinkel, S. N., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160014. https://doi.org/10.1098/rstb.2016.0014

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 65-74. https://doi.org/10.1016/j.biopsycho.2014.11.004

Harrison, O. K., Garfinkel, S. N., Marlow, L., Finnegan, S., Marino, S., Nanz, L., ... & Stephan, K. E. (2020). The Filter Detection Task for measurement of breathing-related interoception and metacognition. BioRxiv.

Harshaw, C. (2015). Interoceptive dysfunction: Toward an integrated framework for understanding somatic and affective disturbance in depression. *Psychological Bulletin*, 141(2), 311-363. https://doi.org/10.1037/a0038101

Harver, A., Katkin, E. S., & Bloch, E. (1993). Signal-detection outcomes on heartbeat and respiratory resistance detection tasks in male and female subjects. *Psychophysiology*, 30(3), 223-230.

Herbert, B. M., & Pollatos, O. (2014). Attenuated interoceptive sensitivity in overweight and obese individuals. *Eating Behaviors*, *15*(3), 445-448. https://doi.org/10.1016/j.eatbeh.2014.06.002

Jenkinson, P. M., Taylor, L., & Laws, K. R. (2018). Self-reported interoceptive deficits in eating disorders: A meta-analysis of studies using the eating disorder inventory. *Journal of Psychosomatic Research*, 110, 38-45. https://doi.org/10.1016/j.jpsychores.2018.04.005

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., ... & Meuret, A. E. (2018). Interoception and mental health: a roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(6), 501-513. https://doi.org/10.1016/j.bpsc.2018.04.007

Khalsa, S. S., & Lapidus, R. C. (2016). Can interoception improve the pragmatic search for biomarkers in psychiatry?. *Frontiers in Psychiatry*, *7*, 121.

https://doi.org/10.3389/fpsyt.2016.00121

Khalsa, S. S., Rudrauf, D., Sandesara, C., Olshansky, B., & Tranel, D. (2009). Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. *International Journal of Psychophysiology, 72*(1), 34-45. https://doi.org/10.1016/j.ijpsycho.2008.08.010

Kline, P. (1999). *The handbook of psychological testing (2nd ed.)*. London: Routledge.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics, 63*(8), 1279-1292. https://doi.org/10.3758/BF03194543

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America, 49*(2B), 467-477. https://doi.org/10.1121/1.1912375

Meissner, K., & Wittmann, M. (2011). Body signals, cardiac awareness, and the perception of time. *Biological Psychology, 86*(3), 289-297. https://doi.org/10.1016/j.biopsycho.2011.01.001

O'Brien, W. H., Reid, G. J., & Jones, K. R. (1998). Differences in heartbeat awareness among males with higher and lower levels of systolic blood pressure. *International Journal of Psychophysiology, 29*(1), 53-63. https://doi.org/10.1016/S0167-8760(98)00004-X

Palser, E. R., Fotopoulou, A., Pellicano, E., & Kilner, J. M. (2018). The link between interoceptive processing and anxiety in children diagnosed with autism spectrum disorder: Extending adult findings into a developmental sample. *Biological Psychology, 136*, 13-21. https://doi.org/10.1016/j.biopsycho.2018.05.003

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483-488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Schroijen, M., Davenport, P., Van den Bergh, O., & Van Diest, I. (in press). Respiratory perception: From sensation to symptom. In J. E. Cotes, J. E., R. Maynard, B. Nemery, P. (Eds.), *Lung function: physiology, measurement and application in medicine, 7th edition*. Oxford: John Wiley & Sons.

Solomon, B. K., Wilson, K. G., Henderson, P. R., Poulin, P. A., Kowal, J., & McKim, D. A. (2015). A breathlessness catastrophizing scale for chronic obstructive pulmonary disease. *Journal of Psychosomatic Research, 79(1),* 62–68. https://doi.org/10.1016/j.jpsychores.2014.11.020

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). Affect 4.0. *Experimental Psychology*. https://doi.org/10.1027/1618-3169/a000005

Van den Bergh, O., Bogaerts, K., & Van Diest, I. (2015). Symptom perception, awareness and interpretation. In *International Encyclopedia of the Social & Behavioral Sciences, 2nd Edition, Vol 23. Oxford: Elsevier* (pp. 866–872). https://doi.org/10.1016/B978-0-08-097086-8.14137-6

Van den Bergh, O., Witthöft, M., Petersen, S., & Brown, R. J. (2017). Symptoms and the body: taking the inferential leap. *Neuroscience & Biobehavioral Reviews*, 74, 185-203. https://doi.org/10.1016/j.neubiorev.2017.01.015

Van Diest, I. (2019). Interoception, conditioning, and fear: The panic threesome. *Psychophysiology*, 56(8), e13421. https://doi.org/10.1111/psyp.13421

von Leupoldt, A., & Dahme, B. (2005). Differentiation between the sensory and affective dimension of dyspnea during resistive load breathing in normal subjects. *Chest*, *128*(5), 3345-3349. https://doi.org/10.1378/chest.128.5.3345

Walentynowicz, M., Witthöft, M., Raes, F., Van Diest, I., & Van den Bergh, O. (2018). Sensory and affective components of symptom perception: A psychometric approach. *Journal of Experimental Psychopathology*, *9*(2), jep-059716. https://doi.org/10.5127/jep.059716

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063

Webster, K. E., & Colrain, I. M. (2000). The relationship between respiratory-related evoked potentials and the perception of inspiratory resistive loads. *Psychophysiology*, *37*(6), 831-841. https://doi.org/10.1111/1469-8986.3760831

Whitehead, W. E., Drescher, V. M., Heiman, P., & Blackwell, B. (1977). Relation of heart rate control to heartbeat perception. *Biofeedback and Self-regulation*, *2*(4), 371-392. https://doi.org/10.1007/BF00998623

Zamariola, G., Maurage, P., Luminet, O., & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations. *Biological Psychology*, *137*, 12-17. https://doi.org/10.1016/j.biopsycho.2018.06.006

Author notes

This study was supported by the Fonds de la Recherche Scientifique (F.R.S – FNRS), Grant # T.0114.18.

The authors thank Olivier Desmedt, Marta Walentynowicz and Omer Van den Bergh for their helpful feedback throughout the preparation of the study and interpretation of the results, and Marlies Mommaerts and Ann-Sofie De Coussemaker for their help with data collection.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Preliminary results of this study were presented at the annual meeting of the International Society for the Advancement of Respiratory Psychophysiology (ISARP), October 2019, Vevey, Switzerland.

Table 1

Average trial number and occlusion length at each reversal point

Reversal #	Downwards going staircase				Upwards going staircase			
	Trial number		Occlusion length		Trial number		Occlusion length	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	16.18	6.89	278.87	48.86	19.55	7.68	137.22	51.43
2	21.56	8.02	313.14	45.57	25.26	8.35	100.88	50.63
3	30.42	8.32	265.10	45.53	34.60	8.95	149.95	46.34
4	36.55	9.12	289.23	45.43	41.01	10.80	123.04	44.82
5	46.48	12.17	262.53	47.42	49.09	12.16	145.10	42.03
6	53.99	13.35	272.99	46.05	56.29	13.50	135.05	40.57

Table 2

Means, standard deviations and Cronbach's alpha of the included tasks and trait questionnaires.

Questionnaire	**Mean**	**SD**	**α**
ROD task (JND)	74.22	37.1	0.70
Auditory discrimination task (JND)	63.99	23.11	0.24
Heartbeat counting task	0.31	0.21	0.82
Trait positive affect (PANAS)	32.66	6.53	0.88
Trait negative affect (PANAS)	18.69	5.85	0.84
Symptom reporting (DLKL)	77.46	15.78	0.90
Dyspnea catastrophizing (BCS)	13.80	10.22	0.92
Difficulty identifying feelings (TAS-20)	16.76	5.41	0.77
Difficulty describing feelings (TAS-20)	13.55	4.57	0.20
Externally-oriented thinking (TAS-20)	17.88	3.82	0.21
Alexithymia total (TAS-20)	48.19	10.37	0.66
IAQ – neutral sensations	31.27	5.63	0.72
IAQ – attention to unpleasant sensation	29.24	4.48	0.44

Table 3

Correlations between questionnaires and ROD task performance, (1) without controlling for any possible confounding variables, (2) while controlling for the average ratio of inspiratory flow right before the first versus the second occlusion during the respiratory occlusion discrimination task, (3) while controlling for performance on the auditory discrimination task, and (4) while controlling for both. P-values are corrected for multiple testing (per column) using the False Discovery Rate (FDR) method.

Questionnaire	**r (1)**	**p (1)**	**R (2)**	**p (2)**	**r (3)**	**p (3)**	**r (4)**	**p (4)**
Trait positive affect (PANAS)	-0.02	0.96	-0.04	0.99	-0.01	0.96	-0.02	0.98
Trait negative affect (PANAS)	0.01	0.96	-0.02	0.99	0.04	0.96	0.00	0.98
Symptom reporting (DLKL)	-0.12	0.96	-0.09	0.99	-0.08	0.96	-0.07	0.98
Dyspnea catastrophizing (BCS)	0.19	0.60	0.16	0.99	0.25	0.17	0.20	0.56
Difficulty identifying feelings (TAS-20)	-0.04	0.96	-0.04	0.99	-0.04	0.96	-0.04	0.98
Difficulty describing feelings (TAS-20)	0.01	0.96	0.03	0.99	0.01	0.96	0.03	0.98
Externally-oriented thinking (TAS-20)	0.02	0.96	0.00	0.99	0.02	0.96	0.01	0.98
Alexithymia total (TAS-20)	-0.01	0.96	-0.01	0.99	-0.01	0.96	0.00	0.98
IAQ – neutral sensations	-0.04	0.96	-0.05	0.99	-0.06	0.96	-0.07	0.98
IAQ – attention to unpleasant sensation	0.09	0.96	0.14	0.99	0.08	0.96	0.14	0.98

Figure 1

Schematic overview of the set-up and the used equipment.

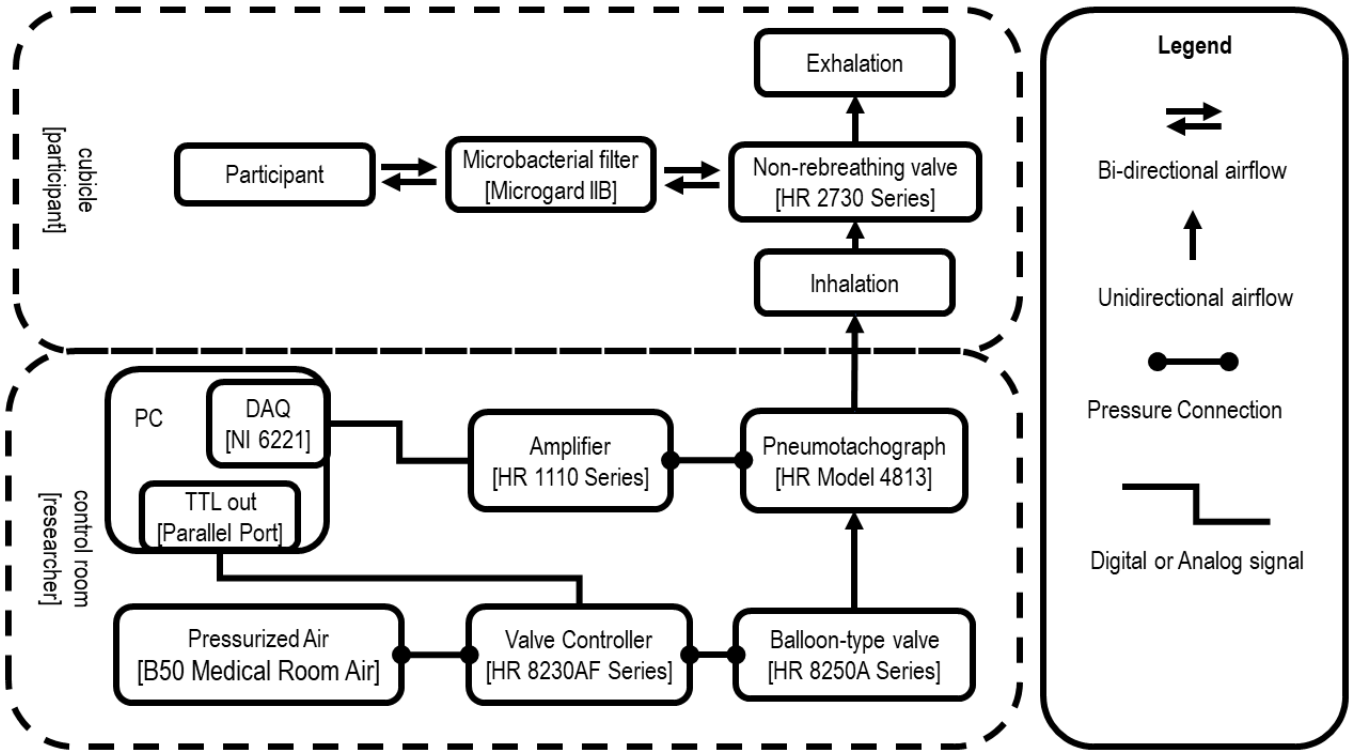


Figure 2

Example of ROD task procedure

