Physiology & Behavior 105 (2012) 982–986

Contents lists available at SciVerse ScienceDirect

Physiology & Behavior

journal homepage: www.elsevier.com/locate/phb

Optimizing reliability and sensitivity of Semmes–Weinstein monofilaments for establishing point tactile thresholds

Elisabeth H. Tracey, Aaron J. Greene *, Richard L. Doty

Smell & Taste Center, Department of Otorhinolaryngology: Head and Neck Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 2 August 2011
Received in revised form 13 October 2011
Accepted 2 November 2011

Keywords:
Tactile
Touch
Threshold
Cutaneous
Sensitivity
Neurology
Diagnostic
Somatosensation
Perception
Methodology
Psychophysics

ABSTRACT

Semmes–Weinstein monofilaments (SWM) are widely used to assess tactile point pressure sensitivity. However, the reliability of SWMs has been questioned, standardization of stimulus presentation procedures is lacking, and the sensitivity measure is commonly confounded by the response criterion. This study sought to assess the reliability of two versions of a forced-choice single staircase SWM test with the goal of optimizing test reliability with a minimum number of test trials. Test–retest and intra-test reliability coefficients for SWM threshold values from the plantar halluces of 24 normal subjects were obtained using two versions of a forced-choice single-staircase procedure. One version followed a two-down one-up rule (2D) and the other a three-down one-up rule (3D). The 3D procedure was significantly more reliable than the 2D procedure for all sequential combinations of reversal pairs. A total of four 3D reversal pairs (i.e., eight reversals) were sufficient to achieve test–retest and intra-test reliability coefficients > 0.90. High reliability with the minimum number of trials was obtained by calculating the threshold as the mean of eight reversals (test–retest $r = 0.93$, $p < 0.001$; Sessions 1 and 2 intra-test $rs = 0.87$ and 0.92; $ps < 0.001$). Identical median detection thresholds were noted for the two repeated test sessions (5.1 g/mm^2). The threshold values correlated with subject age despite the small range of ages tested, suggesting high sensitivity (Sessions 1 and 2 $rs = 0.61$ and 0.63, $ps < 0.001$). This study demonstrates that SWMs provide highly reliable and sensitive point pressure thresholds with very few trials when an appropriate psychophysical paradigm is employed. The brief forced-choice procedure described herein could serve as a basis for standardizing SWM stimulus presentation methods.

1. Introduction

Measurement of tactile point sensitivity is widely employed in neurology and other disciplines to assess peripheral nerve function. Cutaneous pressure, which largely stimulates Merkel and Meissner mechanoreceptors, is conducted by myelinated Aβ fibers, in contrast to pain and temperature sensations, which are mediated by unmyelinated Aδ fibers [1]. Pacinian corpuscles in the hypodermis can also provide tactile information via myelinated fibers, although the degree of force required for their activation is greater than that for the Meissner corpuscles which lie in the dermis. Strong pressure results in an overlapping response from both dermal and hypodermal mechanoreceptors [2,3].

The most popular tests for assessing point-pressure sensitivity are those employing Semmes–Weinstein monofilaments (SWM), the modern version of von Frey hairs. Tests based on SWMs—the only manually administered tests that approach the normal sensitivities of tactile cutaneous receptors [4]—have been criticized for having low reliability, a prerequisite for validity [5,6]. However, reliability information is lacking for most SWM test paradigms, making it difficult to assess the accuracy of any given test procedure. Where such information is available, reliability coefficients vary dramatically. Massy-Westropp [5] assessed SWM sensitivity on fingers and thumbs of seven men and eight women ranging from 27 to 38 years of age two times within a 10-minute period. Poor test–retest reliability was noted for the non-forced-choice 10-trial procedure, with the number of threshold agreements between test and retest being 93 of 180 total tests (52%) and Cohen Kappa values ranging from 0.09 to 0.21. Kaplan-Solms and Saling [7] assessed SWM tactile sensitivity on the left and right breasts of 20 women using three ascending and three descending non-forced-choice method of limits trials. Test–retest reliability coefficients were 0.61 ($p = 0.003$) on the right and 0.36 ($p = 0.06$) on the left. In general, higher reliability coefficients occur for studies that include patients with disorders that depress cutaneous sensitivity, reflecting the influence of distribution variability on the correlation coefficient [8]. For example, a test–retest reliability coefficient of 0.91 was found for the plantar hallux of 25 leprosy patients by Birke et al. using a criterion of two out of three correct responses as the threshold measure [9]. Vinik et al. [10] reliably differentiated 81 diabetics from 32 normal controls using SWMs on two repeated tests of

* Corresponding author at: Smell & Taste Center, 5 Ravdin Building, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA. Tel.: +1 215 662 6580; fax: +1 215 349 5266.
E-mail addresses: etracey@mail.med.upenn.edu (E.H. Tracey), aagreene@sas.upenn.edu (A.J. Greene), doty@mail.med.upenn.edu (R.L. Doty).

0031-9384/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.physbeh.2011.11.002

the dorsal hallux. Agreement for differentiating diabetic from normal subjects was noted on 85.3% of the cases; a Kappa value of 0.88 was reported.

The present study determined, in normal subjects, SWM thresholds and their reliability for forced-choice single staircase threshold procedures employing the two-down one-up rule (2D) and the three-down one-up rule (3D). The goal was to determine which of these procedures was most reliable and the minimal number of trials, i.e., staircase reversals, needed to achieve high reliability and stable threshold measures. Although staircase procedures have been used in some studies of tactile-point pressure sensitivity and tactile spatial acuity [11–13], they are more widely used in other sensory modalities [14–19], given their ability to provide reliable threshold values with minimal numbers of trials. The forced-choice approach results in less variable measures and eliminates the subjects' setting of the response criterion [20].

2. Method

2.1. Subjects

Twenty-four subjects, 12 men and 12 women, were tested. The men ranged in age from 18 to 28 years [mean (SD) = 22 (2.7)] and the women from 18 to 26 years [mean (SD) = 23 (3.7)]. All were in excellent health and were recruited from the staff and student body of the University of Pennsylvania. Written informed consent was obtained in accordance with the requirements of the University's Office of Regulatory Affairs.

2.2. Materials

The Semmes–Weinstein monofilament test kit (SENSELab Aesthesiometer, Hörby, Sweden) was employed. The set consists of 17 flexible nylon fibers of varying diameter and length that are fixed to a plastic handle. The test administrator touches a monofilament to the surface of the skin from a perpendicular angle, applies pressure just until the fiber buckles, and then removes the monofilament. Applied pressures for the monofilaments, which serve as the metric for the dependent measure, range from 1.7 to 137.3 g/mm^2.

2.3. Procedure

The plantar hallux, i.e., the sole of the great toe, was chosen as the test site. This area, which has a relatively homogenous receptor field devoid of hair follicles, is a target site for assessing neuropathies commonly encountered in diabetes [10] and some neurodegenerative diseases, including Parkinson's disease [21]. Importantly, point sensitivity thresholds in this region usually fall in the middle of pressures available from standard SWMs, minimizing basement and ceiling effects that occur on a number of other body sites in normal subjects. Although such effects not problematic for gross clinical assessment of dysfunction, they can preclude accurate determination of subtle deficits, disease progression, and efficacy of clinical interventions.

Each subject was positioned in a chair such that plantar surface of the foot that corresponded to the preferred hand was oriented towards the tester, allowing for perpendicular application of the monofilament to the plantar surface of the hallux. Subjects either elected to cross their leg in a comfortable manner or place their foot on a cushioned surface in front of the chair. A small dot was placed on the center of the plantar hallux with a ballpoint pen to ensure consistent and accurate placement of the monofilament.

Testing was performed on two test occasions separated by an average duration of seven days. During each test session, which lasted ~20 minutes, two different 20-reversal staircase procedures were performed, one after the other in counterbalanced order, as described in detail below. For each procedure, a trial consisted of the presentation of a stimulus and a blank in random order. The subject was required to indicate which of the two stimulus trials seemed to produce the strongest sensation. The subject's eyes were closed during testing and the tester signified the onset of each stimulus by stating, "Here's the first stimulus," and then, "Here's the second stimulus." A response was required even if no sensation was perceived, i.e., the trials were forced-choice. No feedback was provided regarding performance. Responses were recorded manually.

A test session began with the presentation of the monofilament that induced a pressure of 4.5 g/mm^2, followed by or preceded by a blank. If an incorrect response occurred before correct responses were made on five successive stimulus:blank trials at this stimulus level, the next trial was presented with a more stiff monofilament at a level two steps higher. This process was continued until five successive correct responses were made at this or a subsequent stimulus level, which was then defined as the first staircase reversal. If five correct responses were obtained at the initial 4.5 g/mm^2, the next lower stimulus was presented in a similar fashion. In this case, the stimulus level where a miss occurred before reaching the five-trial criterion was considered the first staircase reversal. Once the first reversal occurred, either the two-down one-up rule (2D) or the three-down one-up rule (3D) for movement of the staircase was instituted, depending upon the test procedure to be performed. In the 2D procedure, correct stimulus detection was required on two successive trials before the next lower pressure was presented. In the 3D procedure, three correct successive trials were required. In both instances, a single miss resulted in the presentation of the next stronger monofilament.

Threshold values were computed for every set of sequential reversal pairs for both the 2D and 3D procedures. A threshold was defined as the mean of all the stimulus levels at which a reversal occurred in the included reversal pairs. For two subjects in the 3D procedure and six subjects in the 2D procedure, the lowest stimulus level (1.7 g/mm^2) was reached and the threshold was assigned the conservative value of 1.7 g/mm^2. Test–retest reliability for each set of successive reversal pairs was computed by calculating the Spearman correlation coefficient of the thresholds across the two test sessions. Intra-test reliability was established by computing split-half correlations of alternating reversal pairs with corrections for test length using the Spearman–Brown prophecy formula [22].

3. Results

Median threshold values for the 2D and 3D algorithms, combined across the two test sessions and calculated for cumulative reversal pairs, are presented in Fig. 1. The 2D thresholds were consistently lower, as this algorithm converges on 70.7% correct responses, whereas the 3D procedure converges on 79.4% correct responses [19]. The graph illustrates that the 3D thresholds are stable from the first reversal pair until seven reversal pairs are included in the algorithm, a point where the threshold values increase, possibly reflecting lapses of attention or

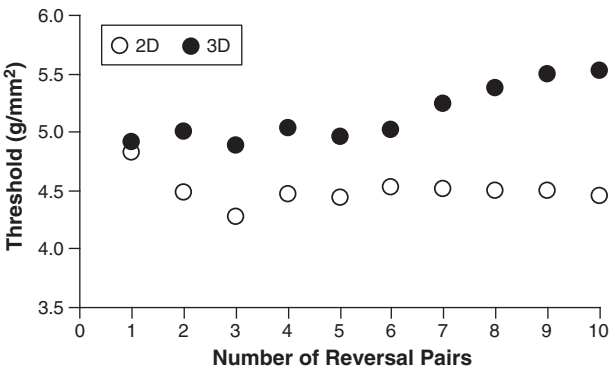

Fig. 1. Median threshold values as a function of the number of cumulative reversal pairs for the 2D and 3D test paradigms.