Lingual tactile acuity, taste perception, and the density and diameter of fungiform papillae in female subjects

Greg K. Essick\textsuperscript{a,b,*}, Anita Chopra\textsuperscript{c}, Steve Guest\textsuperscript{b}, Francis McGlone\textsuperscript{c,d}

\textsuperscript{a}Department of Prosthodontics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7450, USA
\textsuperscript{b}Dental Research Center, University of North Carolina at Chapel Hill, Campus Box 7455, Chapel Hill, NC 27599, USA
\textsuperscript{c}Unilever Research, Port Sunlight, Bebington, Wirral CH63 3JW, UK
\textsuperscript{d}Centre for Cognitive Neuroscience, University of Wales, Bangor LL57 2DG, UK

Received 31 March 2003; received in revised form 4 August 2003; accepted 14 August 2003

Abstract

A growing body of evidence suggests that individuals who differ in taste perception differ in lingual tactile perception. To address this issue, spatial resolution acuity was estimated for 83 young adult females (52 Asians and 31 Caucasians) by their ability to examine with the tongue and identify embossed letters of the alphabet. Ratings of the magnitude of the bitterness of 0.0032M 6-n-propylthiouracil (PROP) were obtained to characterize subjects' taste perception. The density and diameter of fungiform papillae on the anterior tongues of the Asian subjects were measured also. Subjects who rated the bitterness of PROP as very or intensely strong (supertasters) were found to be about 25% more tactually acute than subjects who rated the bitterness as moderate to strong (medium tasters) and twice as acute as subjects who rated it as nondetectable or weak (non-tasters; \( P < .0001 \)): The threshold heights for letter recognition averaged 2.8, 3.5 and 5.4 mm, respectively, for the Asian subjects and 2.6, 3.2, and 5.1 mm for the Caucasian subjects. The thresholds correlated highly with subjects' ratings of bitterness (\( \rho = -0.84 \), \( P < .0001 \)), and for the Asian subjects with the density (\( \rho = -0.84 \), \( P < .0001 \)) and diameter (\( \rho = 0.66 \), \( P < .0001 \)) of fungiform papillae. Mean densities varied from 54.4 cm\(^{-2}\) (non-tasters) to 106.5 cm\(^{-2}\) (medium tasters) to 143.7 cm\(^{-2}\) (supertasters; \( P < .0001 \)). These findings confirm that individuals who differ in taste (PROP) sensitivity also differ in lingual tactile acuity. Tactile and taste sensitivities covary and reflect individual differences in the density and diameter of fungiform papillae on the anterior tongue.

Keywords: Taste; Supertasters; Labelled Magnitude Scale; Psychophysics; PROP; Asian; Fungiform papillae

1. Introduction

Individual differences in taste intensity are, in part, genetically determined and affect food preferences and consumption (see Refs. [1,2] for reviews). Studied most extensively are the differences in individuals' perception of the bitterness of 6-n-propylthiouracil (PROP; and phenylthiocarbamide, PTC, in earlier work). These differences underlie contemporary methods for phenotyping subjects: Supertasters, medium tasters and non-tasters rate the bitterness of PROP as very to intensely strong, moderate to strong, and weak, respectively, and are known to possess corresponding genetic differences. The genetic locus influencing sensitivity to PROP has recently been characterized [3], and the gene that determines PTC sensitivity has been identified [4]. The PTC gene encodes a specific bitter receptor; other related genes may determine the number of chemoreceptor cells (taste buds and fungiform papillae) on the dorsal surface of the anterior tongue, which also determines taste intensity [5,6]. It has been suggested that supertaster status requires both a high density of fungiform papillae, and two dominant alleles of the PROP (or PTC) gene [7].

Survey and laboratory studies conducted over the past decade have identified other differences in the three taster groups. For example, supertasters, medium tasters and non-tasters experience different taste intensities and liking of other bitter, salty, sweet, and fat-containing substances ([2,8–15]; but see Yackinous and Guinard [16]). These differences are of more than academic interest as they...
translate into different lifestyle choices (e.g., liking of fatty foods, tobacco and alcohol) and health risks [1,2,17,18].

Of relevance to the work reported below is the growing body of evidence suggesting that the three groups of tasters also differ in lingual somatosensory function. As for taste intensity, the groups appear to differ in touch intensity. For example, supertasters detect a filament applied to the mid-portion of the tongue with 23 mg-weight of force more frequently than do non-tasters [16]. However, no differences are observed with higher forces of application or on the more sensitive surfaces of the anterior tongue, situations for which all three groups of tasters detect the stimuli equally well. More convincing evidence of a difference in the subjective intensity of somatosensory stimuli has been obtained by studies in which chemicals and foods provided the stimulation. Compared to non-tasters, supertasters experience (i) greater subjective burn and irritancy from capsaicin and other pepperlike substances [19–21], (ii) greater subjective fattiness of fat-containing foods [10], (iii) greater ability to discriminate differences in the fat content of foods [21], and (iv) greater dislike for foods with a high fat content ([11,22]; but see Drewnowski et al. [23]). Although similar distinctions may applied to other aspects of mouthfeel such as thickness and viscosity [14], little is known about the more general aspects of oral texture and form perception. Of relevance to this issue is the finding of Prutkin et al. [2] who found that the two-point discrimination threshold approximates the distance between fungiform papillae on the anterior tongue. However, missing from their preliminary report were data demonstrating that this measure of tactile spatial resolution discriminates subjects who differ in taster status.

In the work reported below, we tested the hypotheses that supertasters have greater spatial resolution acuity on the tongue tip than do medium tasters, who in turn have greater acuity than do non-tasters. We additionally considered the correlation between spatial acuity and ratings of the bitterness of PROP solutions. Based on the previously observed relationship between the two-point discrimination threshold and spacing of fungiform papillae, the extent to which spatial acuity correlates with the density of these taste papillae was also investigated. As a measure of spatial acuity, a novel method recently validated in our clinical laboratory was employed [24]. This method assesses subjects’ ability to identify embossed letters of the alphabet explored with the tongue tip and avoids some of the problems and limitations with measurement of the two-point discrimination threshold.

2. Methods

2.1. Subjects

Eighty-three female subjects between the ages of 18 and 35 years were recruited for testing of lingual tactile acuity. Fifty-two of the subjects, the Asian cohort, were recruited from the Mumbai suburbs, India by Indica Market Research Group (mean age 21 years, \( \sigma = 3.8 \)). The remaining 31 subjects, the Caucasian cohort, were recruited from the staff of Unilever Research, Port Sunlight, UK (mean age 28 years, \( \sigma = 5.7 \)). These cohorts were subsets of larger samples of subjects for whom taster status had been determined. Subjects were randomly selected for recruitment with the constraint that about equal numbers of supertasters, medium tasters, and non-tasters were enrolled. The Indian subjects spoke English fluently and were mostly college students. However, the experimental procedures were explained to them in Urdu, Hindi, or Punjabi in addition to English. Many of the British subjects had participated previously in sensory testing studies at Unilever Research.

Prior to testing, each subject gave informed consent and completed a comprehensive medical questionnaire. Individuals who were pregnant, in poor health, or who reported a history of head injury, ear infections, severe respiratory infections, nasal disorders, dry mouth, thyroid disease or otitis media were excluded [25]. In addition, those taking medications that might affect taste sensitivity were excluded (e.g., oral analgesics such as aspirin or ibuprofen; see Schiffman [26]). The study was approved by the Quality Assurance Committee at Unilever Research, UK, and the Hindustan Lever Research Committee in India.

2.2. Taste intensity of PROP

All subjects were classified as to taster status based on their ratings of the bitterness of 6-n-propyl-2-thiouracil (PROP, Sigma Aldrich, UK). Solutions of three suprathreshold concentrations (0.000032, 0.00032, and 0.0032 M) were prepared in high purity, double deionised water and kept at 4 °C for no longer than 2 weeks before discarding. As needed for testing, aliquots were taken and allowed to equilibrate to room temperature before use. Ten milliliters of each concentration were placed in medicine cups labelled with codes for identification. Subjects were instructed to sip and swish each solution for 10 s while maintaining the head in a downward position. After expectoration, the subjects swallowed the saliva residue and rated the intensity of the bitterness experienced using the Labelled Magnitude Scale (LMS) developed by Green et al. [27]. Subjects entered their responses into a laptop computer by moving a video cursor along a line to which intensity categories were appended (‘no sensation’ to ‘strongest imaginable sensation’). The subjects were instructed to treat the ‘strongest imaginable sensation’ as the most intense sensation they could imagine in any sensory modality. (When this particular definition of the top anchor of the LMS is used, the LMS may be renamed the gLMS, where the ‘g’ denotes ‘general’ [28]). The software extracted and stored the hidden numerical value associated with the location of the cursor along the scale. Subjects rinsed three times with water between solutions. Solutions were randomised and presented a sec-
ond time after each had been delivered once, providing a total of six ratings.

2.3. Taste papillae density and diameter

Measurements of fungiform papillae density and diameter were made on the Asian subjects. After testing of taste sensitivity, the subject rinsed with water and swabbed the anterior dorsal surface of her tongue with blue food colouring (Supercook, UK) using a sterile cotton tip applicator. Vaseline was applied to the lower lip to minimize staining of extraoral tissues. The colouring stained the filiform papillae blue while the fungiform papillae remain unstained [5]. Photographic images were taken of the extended tongue using a Nikon SC-22 camera fitted with a medical lens. The camera lens supported a custom-built deoxidised plate onto which a disposable plastic slide was mounted. On the center of the slide was drawn a square window, 1 cm² in area, using a fine permanent marker. The camera was positioned at the tongue tip with the midline of the tongue bisecting the centre of the window. To ensure that the correct location of the tongue was attained, several images were taken while readjusting the position of the plate slightly between photographs. The clearest image with the desired position was scanned (35 mm LS-2000 film scanner), and the density and size of the fungiform papillae were determined employing in-house image-analysis software. To estimate their size, 24 fungiform papillae were measured within the window. To account for the greater proportion of papillae at the tongue tip, eight were measured from each the right and left distal halves of the window and four from each the right and left proximal halves. Within these regions, papillae were selected randomly for measurement of their diameters. Measurements were made by both an experienced and inexperienced research associate, for which the diameter values differed no more than 0.1 mm for 90% of the subjects.

2.4. Lingual tactile acuity

Spatial resolution acuity on the tongue tip was measured for all subjects. To this end, an up-down tracking procedure was used to estimate the threshold height for the recognition of embossed letters of the alphabet [24]. The 56 stimuli consisted of precision-milled Teflon strips (Shawcross Engravers, UK), one side of which bore an upper case, Arial font, 1-mm-raised letter (A, I, J, L, O, T, U or W) varying in height 2.5, 3, 4, 5, 6, 7, 8 mm. These heights corresponded to font sizes of 10, 12, 18, 21, 28, 30 and 34 points, respectively. The stimuli were autoclaved prior to each experiment at 128 °C for 20 min. After they cooled to room temperature, they were arranged in a grid: seven rows of decreasing letter height (8 mm first row to 2.5 mm,
seventh row) by eight columns according to alphabetic order (A to W).

For testing, the subject was seated comfortably and wore a blindfold to restrict her vision. A letter on the seventh row (the smallest letters) was randomly chosen as the first stimulus and placed in the subject’s hand in the correct orientation. She examined it with the tip of her tongue and identified the stimulus as a letter of the alphabet. Feedback as to the correctness of response was not supplied. An incorrect response led to the selection of a letter one size larger (i.e., from one row higher in the grid) for the next trial; a correct response led to the selection of a letter one size smaller (i.e., from one row lower in the grid). A correct response to the smallest letter or an incorrect response to the largest letter resulted in the selection of a letter of the same size. Letters were selected at random with the constraint that the same letter was not used twice until all other letters of the same height were used. This was achieved by turning used letters over, face down. After all the letters in a given row had been used they were turned face up to be selected again.

The stimulus letter used on each trial and the subject’s response were recorded. A change in performance on any two consecutive trials (correct response followed by an incorrect response, or an incorrect response followed by a correct response) was defined as a reversal. Approximately 40 trials were completed to assure that at least 6–8 reversals in performance were attained [24,29]. Exceptions to this number of trials were made for subjects who consistently identified the smallest letters correctly. The testers did not know the taster status of the Asian subjects. One of the two testers knew the taster status of the Caucasian subjects as she had tested them earlier to obtain this information.

At the end of testing, stimulus letter height was plotted as a function of trial number (Fig. 4). The mean of the letter heights at all reversals was taken as the estimate of the threshold height (horizontal bars in graphs of Fig. 4; also, see Essick et al. [24]). This value estimates the letter height predicted to result in correct identification 50% of the time. Twenty-three subjects (28%) correctly identified the smallest (2.5- and 3-mm-tall) letters on most trials. For these subjects, the threshold was taken as the mean of the stimulus heights. Eight additional subjects correctly identified more than 80% of the time the 2.5-mm-tall letters, which were delivered in over 75% of their trials. For these exceptional subjects, the threshold was set to 2.5 mm.

2.5. Data analysis

Subjects’ consistency in ratings of bitterness to the same concentrations of PROP upon repeated delivery was assessed by characterizing the difference between the second and first ratings ([30]; see Fig. 1). For the Asian cohort, the relationship between perceived bitterness and density of
fungiform papillae on the tongue tip was examined for each concentration (Fig. 3). For this analysis, the two ratings obtained with each concentration were averaged. Mixed-effect model analysis of variance was used to determine whether the relationships differed for the different concentrations. Using the average rating of the highest concentration (0.0032 M) of PROP, each subject was classified as a non-taster, medium taster or supertaster based on cut-off values established by Bartoshuk et al. (personal communication). The classifications of three subjects were changed due to large differences between their two ratings and their comments about the bitterness of the PROP.

For the Asian cohort, the relationship between the density and diameter of fungiform papillae was evaluated graphically and with Pearson’s correlation \( r \) (Fig. 2). Analysis of variance was used to determine whether the fungiform papillae of subjects varying in taster status differed in density and size. Descriptive statistics were calculated to characterize the differences (Table 1). The association between the threshold height for letter recognition and the density and size of the papillae was examined graphically and with Spearman’s correlation \( \rho \) (Fig. 5).

For all subjects, the association between the bitterness of the highest concentration of PROP and the threshold height was examined similarly (Fig. 6). Analysis of variance was used to determine whether the threshold height for letter recognition differed for the three groups of tasters, and if so, whether it differed similarly for the Asian and Caucasian cohorts. To determine whether subjects with above average and below average spatial acuity differed in the letters they recognized, the trial-by-trial data were separated for the two groups. For each letter and height the proportion of trials on which a correct identification was made was calculated and plotted (Fig. 7). Frequencies at which subjects correctly identified each letter for heights within 1 mm of their thresholds were calculated and tabulated (Tables 2 and 3). Errors were examined to identify consistencies in incorrect responses for subjects with above and with below average spatial acuity.

3. Results

3.1. Taste intensity of PROP

The numerical values of bitterness from the LMS varied from 0 to 28 for 0.000032 M PROP, to 86.9 for 0.00032 M PROP, and to 95.1 for 0.0032 PROP. Ninety percent of the second ratings were no more than 9.9 points less and no more than 16.5 points greater than the first ratings, representing a reasonable degree of consistency. As shown in Fig. 1, there was a tendency for the second rating to be systematically higher than the first. For the highest concentration of PROP, the difference was positive (\( P < .0001 \), sign rank test), and Spearman’s correlation was significant for both the highest and intermediate concentrations (\( \rho =.22 \) and .24, respectively; \( P < .05 \)). The difference in ratings in-

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>Non-tasters</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Density of papillae (#/cm(^2))</td>
</tr>
<tr>
<td>Diameter of papillae (mm)</td>
</tr>
<tr>
<td>Threshold letter height (mm), Asian cohort</td>
</tr>
<tr>
<td>Threshold letter height (mm), Caucasian cohort</td>
</tr>
</tbody>
</table>

<p>| Table 2 |
| Errors of subjects with above average spatial acuity |</p>
<table>
<thead>
<tr>
<th>Stimulus letter</th>
<th>% of time identified correctly</th>
<th>Most commonly reported incorrectly as</th>
<th>% of time</th>
<th>Second most commonly reported incorrectly as</th>
<th>% of time</th>
<th>% of time reported correctly or as two most common errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>92</td>
<td>L</td>
<td>4</td>
<td>J</td>
<td>2</td>
<td>98</td>
</tr>
<tr>
<td>O</td>
<td>85</td>
<td>C</td>
<td>4</td>
<td>G</td>
<td>2</td>
<td>91</td>
</tr>
<tr>
<td>T</td>
<td>64</td>
<td>P</td>
<td>9</td>
<td>I</td>
<td>8</td>
<td>81</td>
</tr>
<tr>
<td>L</td>
<td>63</td>
<td>J</td>
<td>15</td>
<td>J</td>
<td>7</td>
<td>85</td>
</tr>
<tr>
<td>J</td>
<td>62</td>
<td>I</td>
<td>27</td>
<td>L</td>
<td>4</td>
<td>93</td>
</tr>
<tr>
<td>U</td>
<td>60</td>
<td>V</td>
<td>22</td>
<td>O</td>
<td>5</td>
<td>87</td>
</tr>
<tr>
<td>A</td>
<td>55</td>
<td>N</td>
<td>8</td>
<td>K</td>
<td>5</td>
<td>68</td>
</tr>
<tr>
<td>W</td>
<td>54</td>
<td>M</td>
<td>9</td>
<td>O</td>
<td>5</td>
<td>68</td>
</tr>
</tbody>
</table>

<p>| Table 3 |
| Errors of subjects with below average spatial acuity |</p>
<table>
<thead>
<tr>
<th>Stimulus letter</th>
<th>% of time identified correctly</th>
<th>Most commonly reported incorrectly as</th>
<th>% of time</th>
<th>Second most commonly reported incorrectly as</th>
<th>% of time</th>
<th>% of time reported correctly or as two most common errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>83</td>
<td>Q</td>
<td>5</td>
<td>C</td>
<td>3</td>
<td>91</td>
</tr>
<tr>
<td>I</td>
<td>82</td>
<td>L</td>
<td>14</td>
<td>J</td>
<td>3</td>
<td>99</td>
</tr>
<tr>
<td>U</td>
<td>55</td>
<td>V</td>
<td>19</td>
<td>O</td>
<td>10</td>
<td>84</td>
</tr>
<tr>
<td>J</td>
<td>50</td>
<td>I</td>
<td>15</td>
<td>V</td>
<td>9</td>
<td>74</td>
</tr>
<tr>
<td>T</td>
<td>49</td>
<td>P</td>
<td>16</td>
<td>F</td>
<td>5</td>
<td>70</td>
</tr>
<tr>
<td>L</td>
<td>33</td>
<td>J</td>
<td>18</td>
<td>C</td>
<td>15</td>
<td>66</td>
</tr>
<tr>
<td>W</td>
<td>30</td>
<td>M</td>
<td>16</td>
<td>U</td>
<td>12</td>
<td>58</td>
</tr>
<tr>
<td>A</td>
<td>18</td>
<td>K</td>
<td>15</td>
<td>O</td>
<td>10</td>
<td>43</td>
</tr>
</tbody>
</table>
creased with their overall magnitude, likely reflecting a context effect [1,31]. That is, solutions were rated more bitter when preceded by solutions that were perceived to be particularly bitter. Based on the average and individual ratings of the highest concentration of PROP, 25 subjects were classified as non-tasters (average rating < 10.9); 27 subjects as medium tasters; and 31 subjects as supertasters (average rating > 61.5; see Fig. 3, right panel).

3.2. Taste papillae density and diameter

Within the Asian cohort, the density of fungiform papillae varied 5.6-fold (from 33 to 184 per cm²) and their diameter, 1.9-fold (from 0.51 to 0.97 mm). As shown in Fig. 2, the diameter decreased linearly with density (τ = −0.83, P < .0001; left panel). The fungiform papillae of the three taster groups differed in both density (P < .0001, middle panel) and diameter (P < .0001, right panel; Table 1).

The anatomical attributes were strongly associated with subject’s ratings of the bitterness of the PROP solutions. As shown in Fig. 3, a monotonically increasing relationship was observed for all three concentrations (from left to right, Spearman’s correlation coefficients τ = .64, .83, and .86; P < .0001). Mixed model analysis of variance additionally confirmed that the slopes of the relationships differed significantly (P < .0001). The slope for the highest concentration of PROP exceeded in value nine times the slope for the lowest concentration, and as such discriminated subjects differing in taster status very well. On average, both supertasters and non-tasters rated the bitterness of the lowest concentration as ‘weak.’ Non-tasters also rated the intermediate and highest concentrations as weak. However, the supertasters rated the bitterness of the intermediate concentration as ‘strong’ and that of the highest concentration from ‘very strong’ to ‘strongest imaginable.’

3.3. Lingual tactile acuity

Thresholds were estimated from 30 to 42 trials (median = 38) providing 2 to 30 (median = 20) reversals in performance during the tracking. Only a few reversals were observed for the most acute subjects who seldom misidentified even the smallest letters. The threshold height for letter recognition ranged from 2.5 to 6.8 mm (mean = 3.7, σ = 1.26).

Data for an Asian non-taster and Asian supertaster are shown in Fig. 4. Note that their thresholds and densities of fungiform papillae differed roughly twofold, suggesting that variation in tactile acuity among subjects might be attributed, in part, to variation in the number of fungiform papillae. Consistent with this prediction, a strong negative correlation was observed for the relationship between these two variables among the Asian subjects (ρ = −.84, P < .0001; left panel of Fig. 5). Given the reciprocal relationship between the density and diameter of the papillae (Fig. 2), a positive correlation was observed between the threshold and diameter (ρ = .66, P < .0001; right panel of Fig. 5).

The above observations predict that a subject’s spatial acuity was associated with her sensitivity to PROP. The left panel of Fig. 6 confirms this relationship. Increased

---

Fig. 3. Monotonically increasing relationship between average ratings of bitterness of PROP and density of fungiform papillae. Test solutions were 0.000032 M (left panel), 0.00032 M (middle panel), and 0.0032 M (right panel). Included are the best fitting regression lines (solid line; r² = .32, .54, and .74, respectively; P < .0001) and the 90% confidence intervals for individual observations (dashed lines). Ratings to the higher concentration were used to classify subjects as non-tasters, medium tasters and supertasters (indicated by different symbols; see text).
bitterness to the highest concentration of PROP strongly correlated with increased acuity ($\rho = -0.84$, $P < .0001$) although the relationship was not linear. Whereas supertasters tended to have the highest acuity, the acuity of non-tasters tended to be more variable with some subjects exhibiting thresholds approximating those of medium tasters and supertasters. Analysis of variance indicated that spatial acuity differed for the three groups of tasters on average ($P < .0001$), and that the differences were the same for the Asian and Caucasian cohorts ($P > .8$). Descriptive statistics are shown in Table 1.

Of interest was whether subjects with higher versus lower acuity recognized all of the stimulus letters equally well. Although the data do not permit a rigorous assessment of this issue, there is evidence that they did not. For example, plotted in Fig. 7 are percentages of correct responses for the letters O, T and A. The left panel represents the pooled data from those subjects whose thresholds were less than the median value and the right panel, from subjects whose thresholds were greater than the median value. As illustrated by the figure, the above average subjects (left panel) could identify all letters at or above threshold levels (50%) at all heights delivered. The below average subjects (right panel) similarly could identify those letters that required only recognition of gross form cues (O and I). In contrast, letters that required detection of their internal structure (A and W) were not identified correctly, in general, by these subjects. Performance on other letters such as T improved with increases in the font size.

To address this issue further, confusion matrix data were tabulated for the two groups of subjects as shown in Table 2 (above average subjects) and Table 3 (below average subjects). Correct and incorrect responses were tabulated only for letters whose heights were within 1 mm of the individual subjects’ thresholds. Consistent with Fig. 7, both groups of subjects were about equally good at identifying I and O at heights within 1 mm of their thresholds (see column 2 of tables). In contrast, the below average subjects could not discern those letters whose identity required appreciation of fine details in their spatial structure.

Further examination of the first and second most common incorrect responses revealed that errors were not random, but made to letters that exhibit spatial features similar to those of the stimulus letter. This was true for both stimuli that were rarely confused and stimuli that were often confused. For example, on the 36% of trials for which T was not identified correctly by the above average subjects, the letter P (Table 2, third column) and the letter I (fifth column) were each reported about one-fourth of the time (9%/36% and 8%/36%; columns 4 and 6). Of the 67% of trials on which the letter L was not identified correctly...

Fig. 4. Threshold tracking data and photographic images of the tongue tips of an Asian non-taster (top) and Asian supertaster (bottom). (Left panels) The stimulus letter height employed is plotted versus trial number. Each minus symbol indicates that the subject identified the letter incorrectly; each plus symbol, correctly. Reversals in performance are identified by circles. Horizontal bars have been drawn at the subjects thresholds, calculated as the mean of the stimulus heights delivered on all reversals in performance. (Right panels) Magnified view of 1 cm$^2$ of epithelium at the tip of the tongue. Fungiform papillae are the pale or pink-stained circular structures surrounded by blue-stained, nongustatory filiform papillae. The non-taster had a threshold of 4.8 mm and a density of 72 fungiform papillae per cm$^2$. The supertaster had a threshold of 2.8 mm and density of 179 papillae per cm$^2$. 

G.K. Essick et al. / Physiology & Behavior 80 (2003) 289–302
295
Fig. 5. Relationships between tactile acuity (y-axes) and the density (x-axis, left panel) and diameter (x-axis, right panel) of fungiform papillae. Included are the best-fitting regression curves (solid line; $r^2=.64$ and $.47$, respectively; $P<.0001$) and the 90% confidence intervals for individual observations (dashed lines). Different symbols identify non-tasters, medium tasters, and supertasters.

Fig. 6. (Left panel) Monotonically decreasing relationship between average ratings of the bitterness of 0.0032 M PROP and the threshold letter height for recognition. Included is the best fitting regression curve (solid line; $r^2=.75$, $P<.0001$) and the 90% confidence intervals for individual observations. Different symbols identify non-tasters, medium tasters, and supertasters. (Middle panel) Mean thresholds for Asian subjects differing in taster status. (Right panel) Mean thresholds for Caucasian subjects. Thin error bars represent $+1$ standard error.
by the below average subjects, the letter J and letter C were reported 27% (18%/67%) and 22% (15%/67%) of the time. These percentages way exceed the 4% predicted for random reporting. Similar results were found for all stimulus letters.

4. Discussion

The results reported above confirm cross-culturally that individuals who differ in the extent to which they perceive the bitterness of PROP also differ in tactile spatial acuity on the tongue (Fig. 6). On average, supertasters were twice as acute as non-tasters (thresholds were half as large in magnitude) and about 25% more acute than medium tasters. This was true for both British Caucasian and Asian Indian subjects. Moreover, acuity on the tip of the tongue correlated highly ($r=.84$) with subjects’ ratings of the bitterness of the concentrated (0.0032 M) PROP solution. Inspection of the scatter plot of the data, however, indicated that many non-tasters exhibited higher thresholds (poorer acuity) than predicted by extrapolation of the data from supertasters and medium tasters to low ratings of bitterness. For subjects for whom papillae were counted (the Asian subjects), acuity was found to correlate highly ($r=.84$) with the density of fungiform papillae on the anterior tongue (Fig. 5). As observed for the correlation with the ratings, thresholds were more variable among non-tasters and less dependent on the density of papillae than for medium tasters and supertasters. As such, the density of fungiform papillae appears to be a less important determinant of lingual acuity of non-tasters than it is of the generally higher acuity of medium tasters and supertasters.

Our findings support and extend the conclusion of Prutkin et al. [2, p. 167] that the “fungiform papillae may act as an array of sensors for detection of oral touch sensations”. The coupling between the density of the array and tactile spatial acuity appears to be greater for medium tasters and supertasters than for non-tasters, who exhibited a greater range of thresholds. Although speculative, the greater variability in thresholds of the non-tasters might reflect differences in strategies for lingual exploration to compensate, in part, for lower densities of fungiform papillae.

A number of additional, albeit indirect, lines of evidence support an anatomical coupling between the fungiform papillae and mechanoreceptive innervation of the anterior tongue. First, as for the density of fungiform papillae, the density of mechanoreceptors supplying the lingual mucosa in human subjects is highest at the tip and the receptive fields are smallest at that location also [32]. Second, fungiform papillae receive both gustatory (chorda tympani nerve) and somatosensory (lingual nerve) afferents. Whereas the gustatory afferents distribute to the taste buds in the papillae, the somatosensory afferents distribute around the taste buds but terminate predominately in the apex of the papilla, an area that is less heavily keratinized and likely

Fig. 7. Proportion of trials (y-axis) on which correct responses were made by subjects who exhibited above average spatial acuity (left) and below average spatial acuity (right) versus stimulus letter height (x-axis). Data plotted for the letters O, T and A.
more susceptible to noxious and mechanical stimulation [33,34]. Third, included among the trigeminal afferents to each papilla (in rats) are 20 to 40 myelinated afferents [35]. The myelinated afferents that supply the lingual mucosa in humans are mechanoreceptors that mediate tactile sensory function [32]. However, their receptive fields in humans are larger than individual papillae and typically exhibit uniform sensitivity rather than punctate sensitivity located at the center of papillae. Fourth, somatosensory differences between taster groups might not extend to areas lacking fungiform papillae. For example, supertasters experience the greatest burn from capsaicin applied to the dorsal surface of the anterior tongue; non-tasters experience the least burn from the greatest burn from capsaicin applied to the dorsal surface of the midportion of the tongue [20]. In contrast, no differences among the three groups are found on the ventral side of the tongue, a surface void of fungiform papillae. Although these observations provide indirect evidence to support a segregation or concentration of the mechanoreceptors with fungiform papillae, additional psychophysical and animal studies are clearly needed to address this issue systematically. As in Karrer et al. [20], psychophysical studies should investigate sensory function at different sites on the tongue, sites which differ and sites which do not differ in the density of fungiform papillae for the three taster groups. Extending future investigations to include circumvallate papillae may also be worthwhile, for this type of papillae is particularly responsive to bitter stimuli [37,38] and so might play a role in determining taster status (although see Mavi and Ceyhan [39]).

4.1. Within- and among-subject variations in the taste intensity of PROP

Ratings of the bitterness of 0.0032 M PROP have been shown useful in classifying subjects into taster groups [1,2,36,40]. Subjects in the present study were also tested with solutions 10 and 100 times weaker in concentration, and they rated each solution a second time after all three had been rated once. Inspection of the data revealed that the second estimates were higher than the first, on average, for the 0.0032 M solution (Fig. 1). A similar trend was suggested by data from subjects who rated the 0.00032 M solution highly, indicating that subjects who experienced intense bitterness during the first round of tasting tended to rate the solutions even higher on the second round. Similar context effects in response to the bitterness of PROP have been reported and have led to underestimation of differences between the three groups of tasters [41]. For example, the saltiness of NaCl is overrated by supertasters when presented after 0.0032 M PROP. Differences in the taste intensities of PROP and NaCl thus appear smaller in magnitude and more similar to those of medium tasters and non-tasters, who do not perceive 0.0032 M PROP as intensely bitter.

Rather than underestimating differences, context effects in the present study appeared to accentuate differences between the three groups of tasters. To illustrate, Fig. 1 shows that averages of the ratings (y-axis) were calculated from pairs for which the second rating was systematically higher than the first (x-axis). This was only true for the higher ratings, those provided mainly by supertasters in response to the concentrated (0.0032 M) solution. As such, average ratings from supertasters differed from those of non-tasters to a greater extent than would have been the case for randomly sampled ratings. As a second illustration consider the right panel of Fig. 3, which shows the linear relationship between the average ratings of bitterness and the density of fungiform papillae. When examined separately for the first and second ratings, the slope of the line is 16% greater in value for the second ratings and an additional 5% of the variability is explained by the regression, indicating greater separation of data points from the different subjects. Although these effects are small, they nevertheless indicate that differences between the groups of tasters were accentuated by context effects.

Excepting the variation contributed by context effects, the magnitude of the differences observed between the two ratings to the same solution (Fig. 1) indicates that an average value provides a more accurate estimate than a single rating. Single ratings can fall on different sides of the cut-off values established for the different taster groups, as was the case for three subjects in this study. In this situation, additional information must be taken into account in order to classify the subjects such as their comments about the taste and facial expressions.

4.2. Among-subject variations in fungiform papillae

Using a modification of the method introduced by Miller and Reedy [5], estimates of the density of fungiform papillae and their diameter were obtained within a 1-cm² square area abutting the tip of the tongue. As in previous studies (see Table 4) we sought to determine the extent to which individuals who differ in taster status differ in the density of fungiform papillae. Given the error inherent in inferring genotype from subjects’ psychophysical ratings [1,3] and the variation expected among subjects within each of the three taster groups [40], we also directly examined the relationship between the ratings and density via scatter plots and correlation. As made evident by Table 4, our findings confirm those of seven previous studies that investigated these issues. Shown in column 2 are the numbers of subjects studied; in column 3, the areas over which papillae were counted; in columns 4, 5 and 6, the mean densities for the three taster groups and their ratios with respect to the mean density of the non-tasters; and in column 7, correlation coefficients and levels of significance for the rating–density relationship. Missing values denote information that was not provided in the publications. In general, all studies found differences among the taster groups with supertasters exhibiting the highest density, and non-tasters, the lowest density. Moreover, ratings of bitterness increased with density sug-
It is also possible that subjects in the present study were recruited nonrandomly to assure about equal numbers of supertasters, medium tasters and non-tasters for testing. However, in contrast to the other two studies, only females were recruited in the present study. Female supertasters have a disproportionately high density of fungiform papillae compared to males and to medium tasters and non-tasters [2,40,42].

It is also possible that subjects in the present study were classified as to taster status more accurately than in Tepper and Nurse [21] and Yackinous and Guinard [16]. These previous studies used a visual–analogue scale and a category scale, respectively, to obtain ratings of PROP and NaCl from subjects. The tastants were presented in random order and subjects were classified by differences between their ratings of bitterness and saltiness. As explained above, the differences might have been compromised (lessened) if NaCl was rated saltier as a result of context effects [41].

Moreover, use of visual–analogue and category scales does not discriminate among subjects who differ in sensitivity as well as the LMS scale used in the present study [1,2,12,13]. All considered, the larger differences in the mean densities observed in the present study might be attributed to enrollment of female subjects only and to the use of scaling methods that optimally discriminate and assign subjects to the different taster groups. It is also likely that context effects in the present study contributed to better separation of subjects who differ in taste intensity; whereas in Tepper and Nurse [21] and Yackinous and Guinard [16], context effects might have resulted in poorer separation of subjects.

Another notable difference in our findings and those of most previous studies is the stronger correlation ($\rho=.86$) between the ratings of bitterness and the density of fungiform papillae (Table 4, column 7). Similar to the greater differences among groups in the mean densities, we attribute the higher correlation to (i) inclusion of females only, (ii) utilization of the LMS to obtain ratings of bitterness, (iii) accentuated ratings from supertasters (context effects), and (iv) counting papillae over a larger area of the tongue, thus affording more accurate estimates of their densities [2,43]. Of particular note are the data from Prutkin et al. [2], which suggest that the correlation exists only for female subjects. As such, inclusion of male subjects as in the other studies would be expected to weaken the observed correlation. We also note that the within-subjects relationship between the number of papillae stimulated and PROP bitterness intensity has been investigated by Delwiche et al. [6]. In this case a strong relationship was found only for PROP tasters. Further studies are needed to clarify gender differences in taste intensity and their relationship to the functional anatomy of the tongue.

In addition to the density of fungiform papillae, the size of the papillae differed among the taster groups as was reported previously by Reedy et al. [36]. The present study extends the earlier finding by demonstrating an inverse (negative slope) linear relationship between the density of papillae and their diameter (Fig. 2). Similar inverse relationships have been demonstrated for the density of mechanoreceptive afferents and the sizes of their receptive fields on the human hand, face and tongue [32].

4.3. Validity of estimates of lingual tactile acuity based on letter recognition

In this study, lingual tactile acuity was estimated by the threshold height for recognition of embossed letters of the tongue.
alphabet examined by the tongue tip. We recently validated this method of acuity and argued that it avoids problems and limitations that are observed with other tests of spatial acuity [24]. For example, lateral scanning and exploratory movements of the tongue are necessary and encouraged for letter recognition, rather than prohibited as during tests of two-point discrimination. Additionally, during tests of two-point discrimination a subject can change her response criteria and use nonspatial information such as the intensity of the stimuli to discern two points of contact from one. In contrast, with letter recognition there are 26 response options (rather than 2) and nonspatial information is less, if at all, useful. That subjects used spatial information in the present study is also attested by the confusion matrix data (Tables 2 and 3). Specifically, common incorrect responses were made to letters that exhibit spatial features similar to those of the stimulus letters as was observed in Essick et al. [24]. This was true for all stimulus letters and for both subjects with above average acuity and with below average acuity. Letters with relatively simple spatial features such as I and O were most easily recognized by both groups, the below average subjects requiring taller letters for recognition. However, letters with complex spatial features such as A and W were relatively more difficult for subjects with below average acuity attesting to the inability of these subjects to discern the internal structure of these letters at any height used in the study.

In a previous study of 20 young adult females (n = 10) and males (n = 10), the threshold height for letter recognition averaged 5.1 mm (σ = 1.1; range 3.7–6.6 mm). In the present study of 83 young adult females, the threshold averaged 3.7 mm (σ = 1.26; range 2.5–6.8 mm). The lower values of the present study cannot be compared, however, to the previous estimates. This is because 37% of the 83 females in the present study were supertasters; whereas, in Essick et al. [24] less than 25% of the subjects were likely to have been supertasters. In Essick et al., [24], female and males subjects were recruited without consideration of taster status, and males, constituting half of the cohort, are less likely to be supertasters than females [40,42]. Moreover, thresholds of the more acute subjects could not be estimated as accurately as in the present study. The letters were made of moulded denture acrylic, and the smallest letters (3-mm-tall) were of marginal quality. In contrast, the letters of the present study were precisely milled in Teflon and the reproduction of even the 2.5-mm-tall letters was excellent.

In this study we delivered approximately 40 trials to assure that eight or more reversals in performance would be achieved for calculation of the thresholds. Similarly to Essick et al. [24], it was found that eight reversals were

![Thresholds estimated from the first eight reversals in performance (y-axis) versus thresholds estimated from all trials delivered (x-axis). The data from four subjects who did not exhibit at least eight reversals are not included. Different symbols identify non-tasters, medium tasters, and supertasters. The line y = x is included on the plot.](image-url)
achieved with fewer trials, i.e., 10–29 trials (median = 15), compared to 11–21 trials (median = 15) in Essick et al. [24].

Also mirroring the results of Essick et al. [24], the estimates from 15 trials were within ± 0.2 mm of the estimates from all the trials for half the subjects and within ± 1.1 mm for 95% of the subjects with no difference in the means ($P > .67$; see Fig. 8). These findings provide additional support for our earlier suggestion that even quickly obtained, small sample estimates of lingual spatial acuity based on letter recognition should be useful in characterizing groups of subjects who differ in spatial acuity [24].

4.4. Implications for food selection

There is evidence to suggest that supertasters differ in the foods they choose to consume [17,18]. Our results indicate that supertasters have generally better lingual tactile acuity than other taster groups. Therefore, it is likely that supertasters will more readily detect small particles and granularity in foodstuffs, in addition to better appreciating bitter tastes (and mixtures of tastants [44]). It seems plausible that supertasters will be more sensitive to food grit and contaminants (and will reject such foods). However, this does not preclude individual supertasters from accepting specific foods that are naturally granular, if the granularity is perceived as desirable. Indeed, the ability to better detect these types of small food ‘features’ does not necessarily translate into a unitary preference for ‘smoother’ foods in the supertaster population. For example, the ability to detect PROP has been shown to be correlated with food preference in different directions for male and female groups. In particular, for males, ratings of PROP bitterness are correlated with a liking for sweets and fats. This trend is reversed for females [45]. Taster status is likely to be only one of many factors that determine food preference. Other influences will certainly include gender, culture and societal factors.

Acknowledgements

We would like to thank Dr. Laurie Lucchina for her valuable assistance with the PROP testing method, and for her helpful feedback concerning the experimental design. Funding for this study was provided by Unilever Research and Development, Port Sunlight, UK.

References

[28] Bartoshuk LM, Duffy VB, Fast K, Green BG, Prutkin J, Snyder DJ. Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: what we have learned from genetic variation in taste. Food Qual Prefer 2002;14:125–38.