

RESEARCH ARTICLE

Developing a trained sensory panel for comparison of different brands of vanilla ice cream using descriptive sensory analysis

C.A. Hettiarachchi* and D.C.K. Illeperuma

Department of Food Science and Technology, Faculty of Agriculture, University of Peradeniya, Peradeniya.

Revised: 07 July 2021; Accepted: 29 August 2021

Abstract: This study aimed at profiling the sensory attributes of vanilla ice cream using descriptive sensory analysis. This method involves both discrimination and quantification of sensory attributes in food products, and required a trained sensory panel. During this study, 26 individuals were screened for normal sensory acuity, and for their capability to identify and quantify different sensory attributes related to vanilla ice cream. Nine individuals conforming to the selection criteria were subjected to a structured training programme. During the programme the individuals were introduced to a range of sensory attributes related to vanilla ice cream, and they were trained on rating the perceived intensities on a line scale in comparison to reference samples that had predetermined attribute intensities. The trained panel was then used to assess three different brands of commercial vanilla ice cream on the selected sensory attributes. The data were statistically analysed using two-way analysis of variance (ANOVA) and mixed model ANOVA. The results showed that yellowness, vanilla flavour, milky flavour, buttery flavour, sweetness and creaminess were significantly different ($p < 0.05$) among the brands, whilst no significant difference ($p < 0.05$) was observed for the intensities of whey flavour and mouth coating character. Identification and quantification of sensory attributes of the available brands provide directional information to a producer to improve the existing formulation and thereby remain competitive in the market.

Keywords: Descriptive sensory analysis, sensory attributes, vanilla ice cream.

INTRODUCTION

Sensory quality of a food product depends on the stimulation of all five or a few sensing organs collectively by its sensory attributes as it is consumed (Rousseau, 2004; Meilgaard *et al.*, 2006). To assess the differences in sensory quality, it is required to quantify these overlapping sensory attributes independently. In

descriptive sensory analysis, sensory attributes related to the interested product class are identified and the intensity of each attribute is then quantified using either a line scale or a category scale (Stone & Sidel, 2004; Meilgaard *et al.*, 2006). The independent assessment of each sensory attribute requires training to overcome the overlapping effects caused by near simultaneous sensation of other sensory attributes that are not evaluated at the same time (Meilgaard *et al.*, 2006).

Sensory skills vary from person to person and some are less sensitive in identifying product differences. It has been reported that ~30 % of any population cannot satisfactorily discriminate among sensory attributes of the products that they regularly consume (Stone & Sidel, 2004). Thus, prior to training it is necessary to select the individuals with normal sensory acuity and who have the ability to distinguish among several sensory attributes. Guidelines for the selection of panellists are given by Stone and Sidel (2004), Meilgaard *et al.* (2006), and in standards such as ASTM Special Technical Publication-758 (1981) and ISO 8586-1 (1993). An appropriate screening procedure results in a set of potential individuals who are capable of developing their skills through a structured training programme (Issanchou *et al.*, 1995).

The method of training primarily depends on the type of descriptive sensory analysis technique adopted for the study. Flavour profile method, texture profile method, quantitative descriptive analysis (QDA) method and spectrum method can be given as examples for different descriptive sensory analysis techniques. More information about these techniques is given by Carpenter *et al.* (2000), Stone and Sidel (2004), Meilgaard *et al.* (2006) and Lawless and Heymann (2010). During

* Corresponding author (charithuop@yahoo.com)

the present study, a combination of QDA method and spectrum method was used for the training of panellists and it consisted of the following steps; introduction of the screened panellists to the target product class (*i.e.* vanilla ice cream), identification of sensory attributes, provision of reference samples for each sensory attribute, rating the perceived intensity for the reference samples, re-evaluation of reference samples and self-adjustment by panellists to the panel mean intensity ratings. In addition, a practice session was carried out using one of the vanilla ice cream brands intended to be rated during the test session.

The sensory quality of an ice cream (or any other food product) plays a major role in determining its consumer acceptance and marketability. Thus, it is worthwhile for the food manufacturers to develop sensory profiles for their products and the competitive products in the market, and these profiles will enable them to identify the key sensory attributes and their relative intensities that have led a particular product to gain a higher market share. Hence, descriptive sensory analysis will inform the manufacturers how their existing products have to be modified in order to match with the preferences of the consumer.

Descriptive sensory analysis of ice cream has been carried out previously to assess the effects of fat replacers on the sensory properties (Ohmes *et al.*, 1998; Roland *et al.*, 1999; Prindiville *et al.*, 2000; Liou & Grün, 2007). In addition, the role of different ingredients on the sensory attributes of ice cream have been assessed by Koeferli *et al.* (1996), Guinard *et al.* (1997), Cody *et al.* (2007), BahramParvar *et al.* (2013) and García-Segovia *et al.* (2013) using descriptive sensory analysis. Moreover, descriptive sensory analysis has been also applied to determine the effects of processing conditions (Inoue *et al.*, 2012) and storage temperature (Buyck *et al.*, 2011) on the sensory attributes of ice cream. Descriptive studies carried out for the comparison of competitive products are not easy to be found due to the proprietary nature of the resulting data. In this study, three popular vanilla ice cream brands in the Sri Lankan market were subjected to descriptive sensory analysis using a trained panel to investigate the variability of key sensory attributes that contribute to their marketability.

METHODS AND MATERIALS

Materials

Materials used for this study are given below in the order they appear in the text. Sugar (Korach Industry Co.,

Ltd., Bangkok, Thailand); citric acid (Weifang Ensign Industry Co., Ltd., Shandong, China); salt (Kotmale Food Products Pvt. Ltd., Kotmale, Sri Lanka); quinine sulfate (JK Trading Co., Colombo 11, Sri Lanka); ferrous sulfate (Lianyungang Kede Chemical Industry Co., Ltd., Jiangsu, China); caramel flavourant (Duckworth Flavours India Pvt. Ltd., Bangalore, India); vanilla flavourant (Star Brand International Flavour and Fragrances India Ltd., Bangalore, India); mint flavourant (Expressions Aromatiques, Mounas-Sartoux, France); dairy character flavourant (Aromco Ltd., Royston, England); lemon flavourant (Expressions Aromatiques), orange flavourant (Expressions Aromatiques); green ginger essence (Ceylon Cold Stores PLC, Ranala, Sri Lanka); cream soda (Ceylon Cold Stores PLC); yellow colourant (Star Brand International Flavour and Fragrances India Ltd.); milk powder (Highland[®], Milco Pvt. Ltd., Colombo 05, Sri Lanka); mayonnaise (Realmayo[®], Kraft Foods Global Inc., Chicago, IL, USA); set yoghurt (Highland[®], Milco Pvt. Ltd.); Austrian processed cheese (Happy cow[®], Gebrüder Woerle GesmbH, Henndorf, Austria); pineapple jam (MD[®], Lanka Canneries Ltd., Colombo 05, Sri Lanka); sterilised milk (Highland[®], Milco Pvt. Ltd.); orange cordial (MD[®], Lanka Canneries Ltd.); condensed milk (Milkmaid[®], Nestlé Lanka PLC, Colombo 10, Sri Lanka); tomato sauce (MD[®], Lanka Canneries Ltd.); whey powder (Crino[®], Agropur Cooperative, Québec, Canada); milk cookies (Little Lion[®], Little Lion Associates (Pvt.) Ltd., Colombo 13, Sri Lanka) butter (Highland[®], Milco Pvt. Ltd.); cream cheese (Philadelphia Regular[®], Kraft Foods Global Inc.); skim milk powder (Anchor[®] non-fat milk powder, Fonterra Limited, Auckland, New Zealand); whip topping (Red Man[®], Phoon Huat and Co., Pte. Ltd., Pandan Loop, Singapore).

Pre-screening of individuals

A pre-screening questionnaire was distributed among 26 undergraduates of the Department of Food Science and Technology, Faculty of Agriculture, University of Peradeniya who had followed a 15 h course on sensory evaluation. Interest and availability, lactose-tolerance, natural dentition, normal health conditions, abstinence from smoking, alcohol and chewing betel, and no dietary restriction on calorie intake were included in the pre-screening questionnaire as major criteria for recruitment.

Subjects conforming to the pre-screening criteria took part in a scaling test, which consisted of 10 figures shaded to different degrees. The subjects were requested to rate the percentage of the shaded area of each figure on a line scale (60 mm in length and anchored at left and right ends, with verbal labels 'None' and 'All'). Subjects

who rated the shaded area of at least 7 figures out of 10 within the tolerance range (correct rating ± 5 mm) on the given line scale were selected for the core-screening programme.

Core-screening of individuals

The core screening programme consisted of 2 matching tests, 4 triangle tests and 5 rating tests. The two matching tests were conducted on the same day, while the triangle tests and rating tests were conducted later on two different days. All the sensory evaluation sessions were carried out adhering to the ethical guidelines of the University of Peradeniya. Presentation of samples during the tests was carried out according to the standard practice (*i.e.* use of 3-digit numbers for labelling, randomisation of samples etc.).

The first matching test was aimed at determining the ability of the individuals to detect the 4 basic tastes. Thus, the first set presented to the individuals composed of 4 samples possessing sweet (sugar, 15 g/L), sour (citric acid, 0.5 g/L), salty (salt, 2 g/L) and bitter (quinine sulfate, 0.05 g/L) tastes. The second set composed of 6 samples, which included the above four samples and a sample with metallic taste (ferrous sulphate, 0.01 g/L) and a sample of distilled water. The samples were prepared by dissolving the stimulant in distilled water and the concentration of each stimulant was above the recognition threshold. The individuals were asked to select the 4 samples in the second set that match with the first set. In addition, they were asked to describe each perception.

In the second matching test, the first set included 4 samples with caramel (caramel flavourant), vanilla (vanilla essence), mint (mint flavourant) and dairy (dairy character flavourant) aromas. The second set composed of 7 samples, which included the above four samples, and three other samples containing lemon (lemon flavourant) orange (orange flavourant) and ginger (green ginger essence) aromas. For the preparation of samples, approximately 100 μ L of the flavourant was absorbed into a swab of cotton wool and it was placed inside a bottle with a volume of 30 mL. The bottles were kept open for 1 h prior to capping to prevent the accumulation of a strong aroma that may cause carryover effects. The individuals were asked to select the 4 samples in the second set that match with the first set. They were instructed to rest for about 3 min in between evaluating the samples by taking several breaths away from the samples. The individuals were also asked to describe the perceived aromas.

A set of triangle tests was then conducted for the individuals to determine their ability to distinguish among different intensity levels of a given sensory attribute. Samples with different intensities of yellowness and sweetness were prepared and 2 triangle tests were conducted for each attribute (Table 1). The individuals were presented with 3 samples at a time, where 2 of them had identical intensities, whilst the other had a different intensity. The individuals were asked to pick the sample with the different intensity. For the rating tests, the individuals were selected based on their abilities to match and describe all the samples in the taste matching test, match all the samples and describe at least 2 out of 4 of the aromas in the aroma matching test and give correct responses for at least 3 out of 4 the triangle tests.

Table 1: Samples and their intensity levels used for the triangle tests

Attribute	Samples used to obtain the two different intensity levels
Yellowness	Cream soda mixed with distilled water (% v/v)
Test 1	35 and 60
Test 2 ^a	40 and 60
Sweetness	Sugar dissolved in distilled water (g/L)
Test 1	11.7 ^b and 0
Test 2 ^a	7.0 ^c and 0

^a The magnitude of difference between the intensities of the two samples provided was always lower in test 2 than in test 1

^b Five times higher than the threshold level

^c Three times higher than the threshold level

Five rating tests were conducted on the sensory attributes, yellowness, milky flavour, sweetness, firmness and thickness. For each attribute, samples representing 4 different intensity levels were used (Table 2). The selected individuals rated one attribute at a time and they received all 4 samples selected for that attribute simultaneously. They were asked to rate the perceived attribute intensity for a given sample on a 150 mm line scale, which was anchored at both left and right ends with verbal labels 'None' and 'Strong', respectively. Re-evaluation of the samples was permitted. The method of evaluation for each sensory attribute was mentioned in the ballot paper.

The individuals to be trained as panellists for the descriptive sensory analysis of vanilla ice cream were selected based on their ability to rate the given samples in the correct order in 4 rating tests out of 5 using a

Table 2: Samples and their intensity levels used for the rating tests

Attribute	Samples used to obtain the four different intensity levels ^a			
Yellowness	Yellow colourant added to milk ($\mu\text{L/L}$)			
	0	100	200	400
Milky flavour	Milk powder dissolved in water (g/L)			
	80	100	130	150
Sweetness	Sugar dissolved in milk (g/L)			
	30	45	60	100
Firmness	Mayonnaise	Set yoghurt	Austrian processed cheese	Pineapple jam
Thickness	Sterilised milk			
		Orange cordial	Condensed milk	Tomato sauce

^a Samples given from left to right are in the order of increasing attribute intensity

considerable length of the given scale, whilst in the test rated incorrectly, only one reversal of ratings between an adjacent pair of samples was allowed. Procedures for both pre-screening and core-screening were adopted from Meilgaard *et al.* (2006).

Final training programme

Four training sessions were held within 2 consecutive weeks for a total time of ~10 h. In the first session of training, an orientation was given to the selected individuals on descriptive sensory analysis. Afterwards, they were provided with one of the vanilla ice cream brands to be evaluated during the test sessions, and asked to identify the perceived sensory attributes of the given sample. A list of sensory attributes related to ice cream together with their definitions obtained from previous studies (Specter & Setser, 1994; Koferli *et al.*, 1996; Guinard *et al.*, 1997; Prindiville *et al.*, 1999; Roland *et al.*, 1999; Lee & Resurreccion, 2002; Yeh *et al.*, 2002; Jinjarak *et al.*, 2006; Cody *et al.*, 2007) were provided to the panellists in order to facilitate the identification of sensory attributes, and thereby to minimise the training time.

The panellists were given the freedom to report any sensory attribute that is not mentioned in the given list and to modify any definition given for a sensory attribute. A discussion was held with the panellists at the end of the session to decide on a list of sensory attributes and their definitions to be used for the intended descriptive study. The ability to identify a particular sensory attribute by all the panellists was considered as the major criterion for the selection of an attribute.

In the second session of training, the panellists were provided with 2 reference samples possessing distinct intensity differences for each sensory attribute that they have selected during the first session. Reference samples previously identified by Koferli *et al.* (1996), Guinard *et al.* (1997) and Cody *et al.* (2007) (or their substitutes) were used for certain attributes, whilst for the others, reference samples were decided by the authors with the consent of the panel. When selecting reference samples, similarity of the method of evaluation of the reference sample to that of ice cream was considered. The panellists were asked to rate the perceived attribute intensities for the given reference samples on a 150 mm line scale, anchored at 12.5 mm from both left and right ends with verbal labels 'Low' and 'High' (Figure 1). Panel mean attribute intensity was then calculated for each reference sample.

At the commencement of the third session, panellists were given a feedback on their individual and panel mean attribute intensity ratings obtained for each reference sample. The panellists whose ratings were not within the range of 'panel mean attribute intensity \pm 10 mm' for a given reference sample were asked to re-evaluate the sample and to adjust their ratings to fall within the tolerance range (Grosso & Resurreccion, 2002; Yeh *et al.*, 2002). The panel mean attribute intensity values were used as the standard attribute intensities for the reference samples.

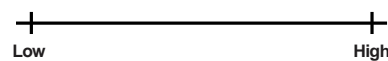


Figure 1: Line scale used for the rating of attribute intensities during the training and test sessions. Not in the actual size

The fourth training session was a practice session in which the panellists were provided with one of the vanilla ice cream brands intended to be used for the test sessions, in white coloured plastic cups with lids and a 5 mL plastic spoon. Ballot papers prepared for each sensory attribute carried two line scales that were similar to the scale used during the evaluation of attribute intensities of the reference samples. One of the line scales was marked with standard attribute intensities (*i.e.* panel mean intensities) of the 2 reference samples identified for that particular attribute, whilst the other line scale was provided to rate the perceived intensity of the practice sample. Reference samples were also made available during the practice session.

The panellists were asked to rate the perceived attribute intensities of the practice sample in comparison to the standard attribute intensities of the reference samples. Panellists were allowed to swallow the samples. They were instructed to rinse their mouth before evaluating a new attribute, and every time when they use a reference sample. Moreover, they were asked to request a new sample for evaluating the textural attributes to minimise the effect of melting due to the exposure of sample to the ambient temperature.

Attribute intensity ratings assigned by the panellists for the practice sample were measured to the nearest millimetre from the leftmost end of the line scale. Standard deviation of the ratings assigned by the panellists for each sensory attribute was then calculated and the attributes with a standard deviation higher than 15 mm were excluded from the test sessions.

Sensory test sessions

During the sensory test sessions, panellists evaluated three brands of vanilla ice cream for the selected sensory attributes. Three sensory test sessions were held to obtain the panellists' ratings in triplicate, and there was an interval of 2 days in between each session. Ice cream samples manufactured within a period of one month before the date of evaluation were purchased from reputed supermarkets (to avoid any contribution of repeated freeze-thaw effects to the sensory quality of ice creams) and stored at -15 ± 2 °C until used. For the presentation of samples, portions of ~30 g were scooped out from the original containers into white coloured plastic cups at least 5 h prior to testing. The cups were closed with their lids and stored at -15 ± 2 °C until they were served to the panellists.

Three brands were served simultaneously to the panellists, and the order of sample presentation for a

given panellist was randomised across the 3 sessions. Prior to each test session, panellists were briefed with the same set of instructions given during the practice session. Ballot papers prepared for each attribute carried 3 line scales (in addition to the line scale with the attribute intensities of the reference samples) to rate the perceived intensities of the 3 samples. The line scales used during the test sessions were identical to those used for rating reference samples and the practice sample. At the end of each test session, attribute intensity ratings assigned by the panellists were measured to the nearest millimetre from the left of the line scale.

Statistical analyses of data

As all panellists evaluated all the samples for a given sensory attribute, randomised complete block design (RCBD) was applied for each attribute. The panellists were considered as blocks, whilst the samples were considered as treatments. Two-way analysis of variance (ANOVA) and mixed model ANOVA were used to find out the significant differences ($p < 0.05$) among the samples for each attribute. Panellist \times sample (*i.e.* block \times treatment) interactions were also estimated as the experiment included repeated judgments. Mean separation was carried out using Fisher's least significant difference (LSD) test. SAS statistical software (v6.12, SAS institute Inc., Cary, USA) was used for data analyses.

RESULTS AND DISCUSSION

Pre-screening and core-screening sessions

Among the 26 undergraduates who received the pre-screening questionnaire, 21 conformed to the recruitment criteria. Seventeen out of the 21 met the expected standards of the scaling test, thus they were selected for the core-screening sessions, which comprised 2 matching tests, 4 triangle tests and 5 rating tests.

Matching tests were aimed at identifying the ability of an individual to discriminate and verbally describe the perceived sensory attributes, when several stimuli are presented together. In the first matching test, all the individuals were able to match sweet, sour, salty and bitter tastes. However, only 15 were able to correctly describe the perceived sensory attribute. In the second matching test, 16 individuals correctly matched caramel, vanilla, mint and dairy aromas and 14 out of those were able to describe at least 2 out of the 4 aromas they matched. None of them were able to correctly describe all the 4 aromas they matched. Terms used for expressing aroma characters such as citrus for lemon, Delta[®] toffee for caramel, cream

for dairy and peppermint for mint were also accepted as correct descriptions. Selection of aromas for the second matching test was based on their relevancy to the target product class. Issanchou and Lesschaeve (1995) have recommended using the target product class during the screening tests, as the sensory ability of an individual may vary according to the product type.

Triangle tests were conducted to identify the ability of the individuals to discriminate between different intensity levels of yellowness and sweetness (Table 1). All the individuals were able to correctly identify the different sample in both test 1 and test 2 held for sweetness. However, in the triangle tests held for yellowness, two individuals were not capable of identifying the different sample in test 2. Two other individuals were unable to distinguish the different sample in test 1, although they gave correct response for test 2, in which there was a lower magnitude of difference between the intensities of the two samples provided (Table 1). Thus, both test 1 and 2 on yellowness were repeated for those two individuals to verify whether they have guessed the answers. During their second attempt, the two individuals were able to correctly identify the different sample in both test 1 and 2. Selection of sensory attributes for the triangle tests were again based on their relevancy to the target product class.

Based on the results of the matching tests and triangle tests, 14 individuals qualified for the rating tests. However, only 12 participated in the rating tests as 2 fell ill. The objective of conducting rating tests was to determine the ability of the selected individuals to rate the perceived intensity differences for a given sensory attribute in a quantitative manner. Four individuals were able to rate the provided samples (Table 2) in the correct order of intensity for all the sensory attributes evaluated. Five other individuals were able to rate the samples in the correct order of intensity for 4 sensory attributes, and there was only one reversal between an adjacent pair of samples in the incorrectly rated attribute. Others were unable to detect the attribute intensity differences up to the expected standards. Therefore, at the end of core-screening sessions, 9 individuals qualified to be trained as panellists for the descriptive sensory analysis of vanilla ice cream.

Training on descriptive sensory analysis

During the first session of training, panellists named the sensory attributes that they can identify in the given vanilla ice cream sample with reference to the list of sensory attributes and their definitions provided. No

new sensory attributes (other than in the given list) were identified by the panellists for the given sample. The panellists preferred to use the same definitions in the given list for all the sensory attributes that they have selected, whilst some definitions were either simplified or modified as proposed by the panel. The sensory attributes and their definitions selected by the panellists at the end of the first training session are given in Table 3. Panel consensus on a common definition for each sensory attribute was considered as an important measure, as it ensures that all the panellists evaluate a given sensory attribute in the same fashion leading to enhance the panel performance during the forthcoming sessions (Munoz & Civille, 1998).

In the second training session, 2 reference samples were provided for each sensory attribute identified during the first training session (Table 4). Panel mean attribute intensities calculated for the reference samples were used as their standard attribute intensities (Lee & Resurreccion, 2002) (Table 4). As one of the panellists inversely rated the reference samples given for slipperiness, he was briefed about the method of evaluation and allowed to

Table 3: List of sensory attributes and their definitions selected during the first training session

Attribute	Definition
Yellowness	The strength of the yellow colour from light to dark yellow ^a
Vanilla flavour	Perception of vanilla flavour ^b
Milky flavour	Perception of the flavour of cooked cow's milk ^c
Whey flavour	Perception of the flavour of the liquid part of milk removed from the curd during the cheese making process ^c
Buttery flavour	Perception of the flavour of butter ^c
Sweetness	The strength of the sweetness from mild to strong ^a
Firmness	The force necessary to compress the sample by the tongue against the roof of the mouth ^d
Slipperiness	The amount in which the sample slide across the tongue ^e
Creaminess	Combination of thickness and lubricative feeling as ice cream melts ^f
Mouth coating	Residual perceived on oral surfaces after sample is swallowed ^g

^a Meilgaard *et al.*, 2006; ^b Cody *et al.*, 2007; ^c Koefler *et al.*, 1996;

^d Specter and Setser, 1994; ^e Lee and Resurreccion, 2002; ^f Prindiville *et al.*, 1999; ^g Yeh *et al.*, 2002

re-assess the two reference samples prior to including his ratings for the calculation of mean attribute intensity for slipperiness. Reference samples were pre-determined by the authors in this study, rather than allowing the panellists to decide, to reduce the training time.

Efforts were taken to prepare the reference samples to closely resemble the target product class. Use of milk as the base to obtain samples possessing different intensities of sweetness and vanilla flavour (instead of using water or propylene glycol as the medium) can be given as an

Table 4: Reference samples provided for each sensory attribute and their standard attribute intensities

Attribute	Reference samples		Calculated standard attribute intensity (mm) ^a
	(A) Low intensity reference	(B) High intensity reference	
Yellowness	(A) Sterilised milk		8
	(B) Sterilised milk with yellow colourant (0.6 mL/L)		130
Vanilla flavour	(A) Vanilla flavourant added to sterilised milk (10 mL/L) ^b		75
	(B) Vanilla flavourant added to sterilised milk (15 mL/L) ^b		121
Milky flavour	(A) Sterilised milk diluted 4-fold with distilled water		38
	(B) Sterilised milk		96
Whey flavour	(A) Whey powder dissolved in water (50 g/L)		52
	(B) Whey powder dissolved in water (100 g/L)		113
Buttery flavour	(A) Milk cookies		46
	(B) Butter		130
Sweetness	(A) Sugar dissolved in sterilised milk (30 g/L)		25
	(B) Sugar dissolved in sterilised milk (120 g/L)		124
Firmness	(A) Mayonnaise		23
	(B) Pineapple jam		114
Slipperiness	(A) Cream cheese		48
	(B) Set yoghurt		120
Creaminess	(A) Skim milk (reconstituted)		27
	(B) Whip topping		113
Mouth coating	(A) Mayonnaise		97
	(B) Cream cheese		125

^a Panellists used a 150 mm line scale to rate the perceived intensities for the reference samples. The given values are the panel mean intensities rounded to the nearest millimetre.

^b Both samples contained yellow colourant (0.4 mL/L) and sugar (80 g/L)

example for such approach. Providing reference samples has been reported to play a key role in increasing the parity among the panellists with regard to identification and quantification of sensory attributes (Drake *et al.*, 2001; Mirarefi *et al.*, 2004). Moreover, to familiarise the panellists with scale usage, the same scale to be employed during the test sessions was used for the rating of reference samples (Meilgaard *et al.*, 2006) (Figure 1).

During the third session, panellists were given a feedback about their individual and panel mean attribute intensity ratings recorded for the reference samples. Table 5 shows the number of panellists who had an individual rating beyond the tolerance range (*i.e.* panel mean attribute intensity ± 10 mm) for a given reference sample.

For the sensory attributes yellowness, vanilla flavour, whey flavour, buttery flavour, sweetness and creaminess, ratings of only 3 or a lesser number of panellists were out of the tolerance range for one or both reference samples (Table 5), revealing that a majority of the panel were able to quantify the relative attribute intensity differences for the given reference samples in a similar manner. Ratings of 4 or more panellists were out of the tolerance range for one or both reference samples provided for milky flavour, firmness, slipperiness and mouth coating. Among them, the low intensity reference sample provided for slipperiness was the most inconsistently rated sample. Such variations amongst the panellists' ratings for a given sensory attribute can occur due to the poor understating of the method of evaluation, or perhaps due to the inability of the provided reference sample to clearly represent the

sensory attribute. All the panellists who had rated attribute intensities beyond the tolerance range re-evaluated the reference samples and adjusted their ratings to the panel mean intensities before the practice session.

The intention of the practice session was to familiarise the panellists with the evaluation of the target product class for the above attribute intensities in comparison to the attribute intensities of the reference samples. Standard deviation (SD) of the ratings assigned for each attribute of the given vanilla ice cream sample revealed that all the panellists were capable of quantifying the attribute intensities in a similar fashion ($SD \leq 15$ mm), except for firmness and slipperiness. Although the panellists were provided with a fresh sample with the intention of minimising the effect of melting on the evaluation of these textural attributes, the received ratings had a $SD > 30$ mm. In a similar fashion, more inconsistencies had been observed by Guinard *et al.* (1997) for the intensity ratings of textural attributes in comparison to flavour attributes. They reasoned this to be a consequence of lack of training and better understanding of flavour concepts than textural attributes by the panel members.

Table 5: Number of panellists rated the attribute intensity of the reference samples beyond the tolerance range

Attribute	Reference samples ^a	No. of panellists rated beyond the tolerance range ^b
Yellowness	(A)	0
	(B)	2
Vanilla flavour	(A)	3
	(B)	3
Milky flavour	(A)	1
	(B)	4
Whey flavour	(A)	1
	(B)	2
Buttery flavour	(A)	3
	(B)	1
Sweetness	(A)	1
	(B)	2
Firmness	(A)	2
	(B)	4
Slipperiness	(A)	7
	(B)	4
Creaminess	(A)	2
	(B)	2
Mouth coating	(A)	4
	(B)	3

^a Refer Table 4 for the names of the reference samples

^b Tolerance range = panel mean intensity \pm 10 mm

Due to the limitation of time, it was decided to exclude these 2 textural attributes from the test sessions, as the panellists needed further training for the quantification of these two attributes. In a typical training programme, a series of practice sessions are conducted until all the panel members show a higher level of consistency in their ratings (*e.g.* $SD \leq 10$ mm; Plemmons & Resurreccion, 1998). Moreover, consistency and reproducibility of the panellists can also be assessed by the comparison of individual ratings received for randomly presented duplicate samples (Guinard *et al.*, 1997). Reliability of the results obtained from a descriptive analysis depends on the level of performance of the panel. Hence, the degree of training required can be decided according to the degree of accuracy required. Moreover, the level of training required also depends on complexity of the product and the number of attributes that have to be assessed during the study (Meilgaard *et al.*, 2006).

Comparison of different brands of vanilla ice cream

Only 7 panellists took part in the test sessions, as 2 of them withdrew themselves due to personal reasons. During the test sessions, each panellist quantitatively evaluated 3 local brands of vanilla ice cream (referred to as S1, S2 and S3 in the text) for 8 sensory attributes in comparison to the standard attribute intensities of the reference samples. The collected data were statistically analysed using two-way ANOVA and the results are given in Table 6. The panellist (*i.e.* block) effect was significant ($p < 0.05$) for all the sensory attributes evaluated showing the variability amongst the panellists in rating the perceived intensities. Such variability is

Table 6: *F* values obtained by two-way ANOVA for the panellist, sample and panellist \times sample interaction effects on the sensory attributes of 3 brands of vanilla ice cream

Attribute	Effect		
	Panellist Degrees of freedom (DF) = 6	Sample DF = 2	Panellist \times sample DF = 12
Yellowness	3.94**	113.95***	0.98
Vanilla flavour	2.78*	21.35***	1.05
Milky flavour	3.04*	49.40***	2.20*
Whey flavour	11.66***	1.81	0.73
Buttery flavour	3.55**	27.62***	5.54***
Sweetness	5.20***	33.28***	1.15
Creaminess	4.99***	27.42***	1.82
Mouth coating	20.75***	5.94**	4.19***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (level of significance)

reported to be common even when using a highly trained panel (Xiong *et al.*, 2002), as each member of the panel tend to use the given scale in different proportions to express the perceived intensity (Meilgaard *et al.*, 2006). Increasing the number of practice sessions and advising the panellists to compare the perceived attribute intensity of the test sample against the intensities of both reference samples (instead of one) before rating the test sample are some of the measures that can be proposed to minimise the panellist effect. In addition, presentation of a warm-up sample (*i.e.* a sample, which had been previously rated by the panel for its attribute intensities) prior to the evaluation of test samples has been recommend by Plemmons and Resurreccion (1998) to improve the panel performance.

The sample (*i.e.* treatment) effect was significant ($p < 0.05$) for all the sensory attributes except for whey flavour, suggesting that the intensity of those attributes differs among the three brands of vanilla ice cream. However, a significant panellist \times sample (*i.e.* block \times treatment) interaction effect ($p < 0.05$) was observed for milky flavour, buttery flavour and mouth coating character, and it indicates that panellists have rated the intensities for each of the above attributes in an inconsistent order across the 3 brands. Thus, it was inconclusive whether the observed treatment effect for the above attributes was solely due to their intensity differences among the brands or due to the inconsistency of the panellists' ratings.

In such circumstances, mixed model ANOVA can be used to determine whether a significant treatment effect exists for those attributes, irrespective of the observed panellist \times sample interaction effect (Stone & Sidel, 2004). An adjusted F value is derived in mixed model ANOVA for this purpose, which is obtained by replacing the residual (or error) mean sum of squares with the panellist \times sample interaction mean sum of squares.

Table 7 presents the adjusted F values obtained for the sensory attributes that showed a significant panellist \times sample interaction effect. The adjusted F values obtained for milky flavour and buttery flavour were higher than the upper α probability value (3.89; Table 7) suggesting that these attributes show a significant difference ($p < 0.05$) in their intensities across the 3 vanilla ice cream brands, although they showed a significant panellist \times sample interaction effect. However, the adjusted F value obtained for the mouth coating character was less than the upper α probability value, thus it was inconclusive whether the intensity of that sensory attribute is significantly different among the 3 vanilla ice cream brands. A similar statistical approach had been followed by Mirarefi *et al.* (2004) to determine whether there is a significant sample effect for attributes that showed a significant panellist \times sample interaction effect, in addition to the sample effect.

Fisher's least significant difference (LSD) test was performed on the sensory attributes that showed a significant difference ($p < 0.05$) among the 3 vanilla ice cream brands to identify which brands were significantly

Table 7: Adjusted F values obtained by mixed model or ANOVA for the sensory attributes with significant panellist \times sample interaction effect

Attribute	Adjusted F value
Milky flavour	22.439
Buttery flavour	4.982
Mouth coating	1.420
Upper α probability value of $F(0.05; 2, 12)^a = 3.89$	

^a F (level of significance: degrees of freedom for the treatments, degrees of freedom for the panellist \times treatment interaction)

Table 8: Mean attribute intensities and Fisher's least significant difference (LSD) values obtained for the sensory attributes that showed a significant treatment effect

Treatment (brand)	Mean attribute intensity (r m)					
	Yellowness	Vanilla flavour	Milky flavour	Buttery flavour	Sweetness	Creaminess
S1	72.43 ^c	72.91 ^b	95.67 ^a	88.10 ^a	88.19 ^b	85.76 ^a
S2	141.19 ^a	106.14 ^a	85.00 ^b	84.52 ^a	110.95 ^a	90.19 ^a
S3	110.10 ^b	75.48 ^b	53.05 ^c	62.71 ^b	74.43 ^c	59.71 ^b
LSD	8.26	11.42	9.01	7.46	9.13	8.98

Means followed with the same letter are not significantly different ($p < 0.05$)

different from each other. Results of the Fisher's LSD test are summarised in Table 8. Intensity of the sensory attributes yellowness, milky flavour and sweetness was found to be significantly different ($p < 0.05$) among the brands. S2 was the sweetest with the most intense yellowness, whilst S1 possessed the most prominent milky flavour. The intensity of buttery flavour and creaminess were similar between S1 and S2, but they were significantly higher in comparison to those recorded for S3. There was no significant difference between S1 and S3 for the vanilla flavour intensity, but it was significantly lower in both brands in comparison to S2.

However, it should not be misunderstood that the marketability of a particular brand depends upon receiving the highest intensity ratings for all or a majority of the attributes. Consumer preference for a particular brand within a given product class depends on its unique proportion of different sensory attribute intensities, which are not necessarily required to be the highest. Conducting descriptive sensory analysis studies in conjugation with consumer preference tests will allow to identify of the preferred levels of intensities for different sensory attributes.

CONCLUSION

Descriptive sensory analysis conducted using an in-house trained panel showed that significant differences ($p > 0.05$) exist among the vanilla ice cream brands S1, S2 and S3 for the attribute intensities of yellowness, vanilla flavour, milky flavour, buttery flavour, sweetness and creaminess. No significant difference ($p > 0.05$) was observed among these brands for the attribute intensities of whey flavour and mouth coating character. Further training is required for the identification and quantification of textural attributes of vanilla ice cream. We believe that this work will communicate a systematic approach for the development of an in-house trained sensory panel, and provide a guideline for the establishment of product-specific sensory profiles, which can be used for either new product development or improvement of an existing product.

Acknowledgement

The authors wish to thank all the panellists for their voluntary participation and the Department of Quality Assurance, Ceylon Cold Stores PLC, Ranala, Sri Lanka for providing some of the required materials for this study.

REFERENCES

1. American Society for Testing and Materials (ASTM) (1981). *ASTM Special Technical Publication-758. Guidelines for the Selection and Training of Sensory Panel Members*. American Society for Testing and Materials, Philadelphia, USA.
2. Bahram Parvar M., Tehrani M.M. & Razavi S.M.A. (2013). Effects of a novel stabilizer blend and presence of κ -carrageenan on some properties of vanilla ice cream during storage. *Food Bioscience* **3**: 10 – 18. DOI: <http://dx.doi.org/10.1016/j.fbio.2013.05.001>
3. Buyck J.R., Baer R.J. & Choi J. (2011). Effect of storage temperature on quality of light and full-fat ice cream. *Journal of Dairy Science* **94**: 2213 – 2219. DOI: <http://dx.doi.org/10.3168/jds.2010-3897>
4. Carpenter R.P., Lyon D.H. & Hasdell T.A. (2000). *Guidelines for Sensory Analysis in Food Product Development and Quality Control*, 2nd edition, pp. 46 – 49. Aspen Publishers, Gaithersburg, USA. DOI: <http://dx.doi.org/10.1007/978-1-4615-4447-0>
5. Cody T.L., Olabi A., Pettingell A.G., Tong P.S. & Walker J.H. (2007). Evaluation of rice flour for use in vanilla ice cream. *Journal of Dairy Science* **90**: 4575 – 4585. DOI: <http://dx.doi.org/10.3168/jds.2006-531>
6. Drake M.A., McIngvale S.C., Gerard P.D., Cadwallader K.R. & Cville G.V. (2001). Development of a descriptive language for cheddar cheese. *Journal of Food Science* **66**: 1422 – 1427. DOI: <http://dx.doi.org/10.1111/j.1365-2621.2001.tb15225.x>
7. García-Segovia P., Iborra-Bernad C., Andrés-Bello A., González-Carrascosa R., Barreto-Palacios V., Bretón-Prats J. & Martínez-Monzó J. (2013). Replacing sugar in ice Cream: Fruit Up[®] as a substitute. *Journal of Culinary Science and Technology* **11**: 155 – 164. DOI: <http://dx.doi.org/10.1080/15428052.2013.769865>
8. Grosso N.R. & Resurreccion A.V.A. (2002). Predicting consumer acceptance ratings of cracker-coated and roasted peanuts from descriptive analysis and hexanal measurements. *Journal of Food Science* **67**: 1530 – 1537. DOI: <http://dx.doi.org/10.1111/j.1365-2621.2002.tb10317.x>
9. Guinard J.X., Zoumas-Morse C., Mori L., Uatoni B., Panyam D. & Kilana A. (1997). Sugar and fat effects on sensory properties of ice cream. *Journal of Food Science* **62**: 1087 – 1094. DOI: <http://dx.doi.org/10.1111/j.1365-2621.1997.tb15044.x>
10. Inoue K., Ishii Y., Iwata M., Taketsuka M., Ichihashi N., Keiji I. & Toko K. (2012). Effects of manufacturing process conditions on sensory attributes and microstructure of ice cream. *Sensors and Materials* **24**: 245 – 260
11. International Organization for Standardization (1993). *ISO 8586-1. Sensory Analysis - General Guidance for the Selection, Training, and Monitoring of Assessors - Part I: Selected Assessors*. International Organization for Standardization, Geneva, Switzerland.

12. Issanchou S., Lesschaeve I. & Köster E.P. (1995). Screening individual ability to perform descriptive analysis of food products: basic statements and application to a camembert cheese descriptive panel. *Journal of Sensory Studies* **10**: 349 – 368.
DOI: <http://dx.doi.org/10.1111/j.1745-459X.1995.tb00025.x>
13. Jinjarak S., Olabi A., Jiménez-Flores R., Sodini I. & Walker J.H. (2006). Sensory evaluation of whey and sweet cream buttermilk. *Journal of Dairy Science* **89**: 2441 – 2450
14. Koeflerli C.R.S., Piccinali P. & Sigrist S. (1996). The influence of fat, sugar and non-fat milk solids on selected taste flavor and texture parameters of a vanilla ice cream. *Food Quality and Preference* **7**: 69 – 79.
15. Lawless H.T. & Heymann H. (2010). *Sensory Evaluation of Food: Principles and Practice*, 2nd edition, pp. 231 – 244. Springer Science and Business Media, New York, USA.
DOI: <http://dx.doi.org/10.1007/978-1-4419-6488-5>
16. Lee C.M. & Resurreccion A.V.A. (2002). Improved correlation between sensory and instrumental measurement of peanut butter texture. *Journal of Food Science* **67**: 1939 – 1949.
DOI: <http://dx.doi.org/10.1111/j.1365-2621.2002.tb08750.x>
17. Liou B.K. & Grün I.U. (2007). Effect of fat level on the perception of five flavor chemicals in ice cream with or without fat mimetics by using a descriptive test. *Journal of Food Science* **72**: S595 – S604.
DOI: <http://dx.doi.org/10.1111/j.1750-3841.2007.00494.x>
18. Meilgaard M., Civille G.V. & Carr B.T. (2006). *Sensory Evaluation Techniques*, 4th edition. CRC Press, Boca Raton, USA.
19. Mirarefi S., Menke S.D. & Lee S.Y. (2004). Sensory profiling of Chardonnay wine by descriptive analysis. *Journal of Food Science* **69**: S211 – S217
DOI: <http://dx.doi.org/10.1111/j.1365-2621.2004.tb11007.x>
20. Munoz A.M. & Civille G.V. (1998). Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies* **13**: 57 – 75
DOI: <http://dx.doi.org/10.1111/j.1745-459X.1998.tb00075.x>
21. Ohmes R.L., Marshall R.T. & Heymann H. (1998). Sensory and physical properties of ice creams containing milk fat or fat replacers. *Journal of Dairy Science* **81**: 1222 – 1228.
22. Plemmons L.E. & Resurreccion A.V.A. (1998). A warm-up sample improves reliability of responses in descriptive analysis. *Journal of Sensory Studies* **13**: 359 – 376.
DOI: <http://dx.doi.org/10.1111/j.1745-459X.1998.tb00095.x>
23. Prindiville E.A., Marshall R.T. & Heymann H. (1999). Effect of milk fat on the sensory properties of the chocolate ice cream. *Journal of Dairy Science* **82**: 1425 – 1432.
24. Prindiville E.A., Marshall R.T. & Heymann H. (2000). Effect of milk fat, cocoa butter, and whey protein fat replacers on the sensory properties of low fat and non-fat chocolate ice cream. *Journal of Dairy Science* **83**: 2216 – 2223.
25. Roland A.M., Phillips L.G. & Boor K.J. (1999). Effects of fat replacers on the sensory properties, color, melting, and hardness of ice cream. *Journal of Dairy Science* **82**: 2094 – 2100.
26. Rousseau B. (2004). Sensory evaluation techniques. *Handbook of Food Analysis* (ed. L.M.L. Nollet), 2nd edition, pp. 21 – 22. Marcel Dekker, New York, USA.
27. Specter S.E. & Setser C.S. (1994). Sensory and physical properties of a reduced calorie frozen dessert system made with milk fat and sucrose substitutes. *Journal of Dairy Science* **77**: 708 – 717.
28. Stone H. & Sidel J.L. (2004). *Sensory Evaluation Practices*, 3rd edition. Elsevier Academic Press, San Diego, USA.
29. Xiong R., Meullenet J.F., Hankins J.A. & Chung W.K. (2002). Relationship between sensory and instrumental hardness of commercial cheeses. *Journal of Food Science* **67**: 877 – 883.
DOI: <http://dx.doi.org/10.1111/j.1365-2621.2002.tb10693.x>
30. Yeh J.Y., Resurreccion A.V.A., Phillips R.D. & Hung Y.C. (2002). Overall acceptability and sensory profiles of peanut spreads fortified with protein, vitamins and minerals. *Journal of Food Science* **67**: 1979 – 1985.
DOI: <http://dx.doi.org/10.1111/j.1365-2621.2002.tb08756.x>