An Interlaboratory Study of Firmness, Aroma, and Taste of Pectin Gels

Birgit Lundgren, Rose-Marie Pangborn, Nicole Daget, Masaaki Yoshida, David G. Laing, Robert L. McBride, Nerys Griffiths, Lea Hyvönen, François Sauvageot, Klaus Paulus and Nina Barylko-Pikielna

Birgit Lundgren: SIK - The Swedish Food Institute, P.O. Box 5401 S–402 49 Göteborg (Sweden)
Rose-Marie Pangborn: Department of Food Science and Technology, University of California, Davis, Calif. 95616 (USA)
Nicole Daget: Nestlé Products Technical Assistance Co. Ltd. CH–1814, La Tour-de-Peilz (Switzerland)
Masaaki Yoshida: College of Science and Engineering, Chuo University, Kasuga 1–13–27, Bunkyo-ku, Tokyo (Japan)
David G. Laing and Robert L. McBride: CSIRO, Division of Food Research, P.O. Box 52, North Ryde N.S.W. 2113 (Australia)
Nerys Griffiths: Agricultural Research Council, Food Research Institute, Colney Lane, Norwich NOR 70F (England)
Lea Hyvönen: Department of Food Chemistry and Technology, University of Helsinki, SF–00710 Helsinki 71 (Finland)
François Sauvageot: ENSBANA, Laboratoire de Biologie Physico-Chimique, Campus Universitaire, F–21000 Dijon (France)
Klaus Paulus: Bundesforschungsanstalt für Ernährung, D–75 Karlsruhe, Engesserstrasse 20 (Germany)
Nina Barylko-Pikielna: Food and Nutrition Institute, Powsinska 61/63, Warsaw (Poland)
(Received April 1, 1985; Accepted April 17, 1985)

Subjects at laboratories in ten countries (Australia, England, Finland, France, Germany, Japan, Poland, Sweden, Switzerland and USA) evaluated firmness, aroma, flavor, and taste intensity of pectin gels containing different amounts of pectin and flavoring. With few exceptions, the ten laboratories (16 to 34 subjects at each laboratory; 252 total) reached the same conclusions regarding the sample differences. Also, the relative frequencies of discriminators were similar: All subjects discriminated firmness, and about half of them discriminated other attributes. Each laboratory's data showed significant differences among pectin levels as well as flavoring levels. The average perceived intensities of aroma, flavor, sweetness and sourness decreased with increased pectin level, while firmness increased.

Introduction

Sensory scientists are concerned with the generalizability of results across different laboratories. During the last decade, several interlaboratory sensory studies have been published (1, 2, 3, 4, 5, 6, 7), all with the main purpose of comparing results obtained at different locations. However, since they tested different attributes (appearance, taste, odor or texture), or used different psychophysical methods (differences, intensities, ratio evaluations, or hedonics), or different commodities (foods or non-foods) it is virtually impossible to compare the conclusions from one study to those of another. Nonetheless, these cross comparisons give important information on the kind and degree of variation one can expect among the results when the same experiment is executed at different laboratories, *i. e.,* by different experimenters using different groups of subjects.

Several of the present authors have participated in previous interlaboratory studies on taste discrimination for sodium chloride in water (1), sucrose in both water and orange juice (2), and sucrose in a coffee beverage (3). The three reports analyzed both within and between laboratory variation in taste sensitivity, and degree of liking of simple taste compounds in water or beverages. With only minor discrepancies, the results showed high agreement among different laboratories. The primary purpose of the present study was to expand the number of participating laboratories and to extend the experiments to include yet another psychophysical method, namely intensity measurements, and a more complex model food taking both texture and flavor into account.

Experimental

Subjects

The ten laboratories represented by the authors of this paper participated. They will be referred to by easily comprehensible abbreviations and are listed alphabetically, by order of the names of the participating countries (cf., e.g., **Tab. 4**). A total of 252 people took part, approximately half of each gender. As shown parenthetically in **Tab. 4,** the number of subjects at each laboratory varied from 16 to 34. Subjects were students and/or employees, ranging in age from 18 to 40 years. Most had no previous experience in sensory evaluation.

Experimental plan

A detailed plan, written in English and elaborated jointly by the American, Swedish and Swiss laboratories, was agreed upon by all participants before starting the experiments. The plan contained instructions to the experimenters, the laboratory technicians, and the subjects. The instructions and forms intended for the subjects were translated into the languages spoken at the Finnish, French, German, Japanese, Polish, Swedish, and Swiss laboratories.

Materials

A cube-shaped (side *ca* 2.0 cm; weight *ca* 9 g) orange-flavored pectin gel product was used. Samples of different formulations were made on one occasion by Nestlé Products Technical Assistance Co. Ltd, La Tour-de-Peilz, Switzerland, and distributed to all participants. They consisted of the nine combinations of three pectin levels (1.48, 2.04 and

Tab. 1 Ingredients of the pectin gel samples

Ingredient	Amount (g)		
	(soft)	(medium)	(hard)
Cold water	1750	1750	1750
Icing sugar	400	400	400
Sucrose	2100	2100	2100
Glucose (42 DE)	1900	1900	1900
Citric acid 1:1 (w/v)	80	80	80
Pectin			
(Ruben Jaune; Obi Pectin)	80	110	150
% of ingredient	1.26	1.73	2.35
% of finished product	1.48	2.04	2.77

Flavoring: Orange «soft aroma» (Givaudan)
To the amount above was added:
1.80 ml = 0 280 ⁒ (= high flavoring level)
1.20 ml = 0 190 ⁒ (= medium flavoring level)
0.60 ml = 0 095 ⁒ (= low flavoring level)

2.77% w/w) and three flavoring levels (0.0095, 0.0190 and 0.0280% w/w). A list of ingredients of the product is given in **Tab. 1.** The product was produced in the following way.
A mixture of icing sugar and pectin was boiled in water for one minute. Sucrose and glucose were added, and the mixture was heated to a temperature of 105°C. When the °Brix measured on a refractometer was 77.5, the heating was stopped, and the flavoring agent and citric acid solution were added. (This preparation took about six minutes.) The mass was poured into starch-coated plaster molds with 10 g/cube. The top surfaces of the cubes were powdered with starch and left for 24 hours to cool, then blown to remove the starch. The cubes were spray-coated with a thin layer of liquid paraffin to prevent them from sticking to each other. They were left to dry for another 24 hours and finally packed in small plastic boxes (24 cubes/box), which were sealed with tape and sent to all the laboratories. Each box contained only one layer of samples, placed between corrugated paper at the bottom and silicone paper on top. The samples were separated by plastic dividers to prevent sticking. There were small visual differences among the samples, as they became slightly darker with an increased amount of pectin. The evaluations took place within 2–4 months after the samples had been received.

Instrumental texture analyses
To establish that there were measurable "hardness" differences among the three pectin levels, and no systematic differences among the batches across flavoring levels, breaking force was determined by the Swedish laboratory using an Instron Universal Testing Machine. Using a compression cell, one gel cube was compressed by a stainless steel plunger (crosshead speed: 10 mm/min) with a considerably larger surface (plunger diameter: 60 mm) than the cube (side: 2.0 cm). The pressure was maintained until the cube cracked, which occurred when the cube was deformed to about 50% of its original height. The force of deformation (F) at this breaking point was registered. Measurements were made on ten cubes from different boxes. The average F value was calculated and plotted against the amount of pectin. This plot showed distinct differences in hardness for the three pectin levels (F = 3.1, 5.9, and 10.0, respectively). For each pectin level, the three values for hardness for the different flavoring levels were close and showed no systematic relationship with flavoring level.

Sensory analyses

All subjects performed the following experiments, A–C, and some laboratories also included the optional experiment D (Fin., Fra., Ger., Swi., Swe.).

Experiment A: Aroma intensity (attribute 1)
Experiment B: Firmness (attribute 2)
Experiment C: Total flavor intensity (by mouth) and sweetness intensity (attributes 3 and 4)
Experiment D: Total flavor intensity (by mouth) and sourness (attributes 3 and 5). Optional.

The translations of the attribute names from English are given in **Tab. 2.** Due to a mistranslation of "flavor intensity" as "degree of liking", no data of Experiment C were collected at the Japanese laboratory.
Experiments A–D were performed in that order on different occasions with different sets of samples to avoid halo effects among evaluations of different attributes, and to permit testing of all nine samples on the same occasion. Since experiments A–D required as many as twenty sessions per subject (4 introductory + 4 experimental × 4 replicate), an option was given to exclude experiment D. To become familiar with the technique and samples, each subject took part in 1–2 introductory sessions for each experiment. The results of

Tab. 2 Translations used of the sensory attribute names

Language	English term used in experimental plan[1]				
	aroma intensity	firmness	total flavor intensity	sweetness	sourness
Finnish (lab. 3)	tuoksun voimakkuus	kiinteys	kokonaismaitto	makeus	happamuus
French (labs. 4 and 9)	intensité d'arôme	fermeté	intensité globale de la flaveur	sucrosité	acidité
German (lab. 5)	Gesamt-Aroma-Intensität	Festigkeit	Gesamt-Geschmacks-Intensität	Süsse	Säure
Japanese (lab. 6)			-[2]	-[2]	-[2]
Polish (lab. 7)					-[2]
Swedish (lab. 8)	aromstyrka	fasthet	smakstyrka	söthet	surhet

1) English was used by the Australian, English, and American laboratories (1, 2 and 10)
2) Not performed by this laboratory (cf. text)

these sessions were excluded from the final calculations. Each subject performed a total of four replicates of all experiments, completing all evaluations of one experiment before commencing the next.

Sample preparation and sensory evaluations

Experiment A. Aroma intensity
Sample preparation. A sample consisted of one cube that was cut into four cubes of equal size. The four pieces were placed in a 200–250 ml glass with an opening diameter of approx. 5 cm. The glass was opaque, or covered by aluminium foil, to prevent subjects from seeing the differences in appearance among the samples. The glass was covered by a watch glass and was left for at least 30 min at room temperature (20–24 °C) for equilibration before the aroma evaluation took place.
Sensory evaluation. The subjects were instructed to:
"Leave the glass on the bench. Remove the watch glass and sniff the sample. Judge the total aroma intensity using the scale (a 10 cm line):

none — very strong

Indicate your judgement by a vertical mark on the line. Please use the same sniffing technique each time. Two or three short sniffs are probably better than one long sniff. Replace the watch glass and proceed to the next sample".

Experiment B. Firmness
Sample preparation. A sample consisted of one cube, presented to the subject together with a plate and a serrated knife. (Samples of suitable knives were sent to all laboratories in order for the tools to be as similar as possible in all evaluations).
Sensory evaluation. The subjects were instructed to:
"Place the cube on the plate. Cut the sample into two halves, chew one half 2–3 times and evaluate its firmness using the scale (a 10 cm line):

very soft — very firm

Indicate your judgement by a vertical mark on the line. Note that the knife gets sticky of the samples. Please use the tissue to wipe it off before proceeding to the next sample. (Some laboratories provided as many knives as samples.). Rinse your mouth carefully before evaluating the next sample. Spit out all sample residue and water".

Experiment C: Total flavor intensity (I) and sweetness
Sample preparation.
A sample consisted of one cube. The cube was presented to the subject together with a plate and a serrated knife.
Sensory evaluation. The subjects were instructed to:
"Cut the cube into two halves, and then cut out a slice of one half. (This provided a slice with two freshly cut surfaces and a minimum of paraffin-covered surface).
Place the slice on the middle of your tongue, move it around slowly, chew it 1–2 times, and evaluate the two attributes: total flavor intensity and sweetness intensity using the scales (lines of 10 cm):

none — very strong

Rinse your mouth carefully before evaluating the next sample. Spit out all sample residue and water".

Experiment D. Total flavor intensity (II) and sourness
Instructions were the same as for C, replacing "sweetness intensity" by "sourness intensity".

Statistical evaluations
All evaluations were transformed into scores by measuring the length (in cm) from the left end of the 10 cm scale to the mark made by the subject. For each attribute, the data were subjected to analyses of variance (ANOVAs) in the three different ways described below, corresponding to zero hypotheses about averages on three different levels: 1) individual, 2) one laboratory at a time, and 3) all laboratories simultaneously. Since an ANOVA only gives information on whether or not there are differences among averages, but no information on how the averages are related to the independent variables (in this case flavoring and pectin content, respectively), regression plots were drawn to show these relations.

1. For each subject (= individual data)
Averages were calculated and denoted according to the following:
Average for a sample = average of 4 values = 4 reps.
Total average for a pectin level = average of 12 values = 3 flavoring levels × 4 reps.
Total average for a flavoring level = average of 12 values = 3 pectin levels × 4 reps.
For each subject, a balanced two-way ANOVA was performed with the two variables: 1) pectin content (three levels), and 2) flavoring content (three levels), and their interaction. The main effects and interaction were always tested against the error term provided by the variation among replicates. This ANOVA provided information on the existence of differences among the total averages for pectin and flavoring levels, respectively. The information was considered a measure of a subject's ability to discriminate among samples, and was used to classify that person as a "discriminator" or "non-discriminator" for the attribute in question. A subject was considered a discriminator among pectin levels (flavoring levels) if a significant difference ($p<5\%$) was shown for variable 1 (2) in the mentioned ANOVA.
An "individual regression plot" was drawn using the above-mentioned sample averages. It was inspected visually, and the appearance of the plot ("response pattern") was compared with that of other subjects within the same laboratory and with subjects among all laboratories.

2. For each laboratory (pooled data)
Averages were calculated and denoted according to the following:
Pooled average for a sample = average of (64–136) values = (16–34) subjects × 4 reps.
Grand mean for a pectin level = average of (192–408) values = (16–34) subjects × 3 flavoring levels × 4 reps.
Grand mean for a flavoring level = average of (192–408) values = (16–34) subjects × 3 pectin levels × 4 reps.
Grand mean for a subject = average of 36 values = 3 pectin levels × 3 flavoring levels × 4 reps.
For each laboratory, a balanced three-way ANOVA was performed with the three variables: 1) pectin content (three levels), 2) flavoring content (three levels), 3) subjects (16–34 levels), and all interactions (1×2; 1×3; 2×3; and 1×2×3). The main effects and interactions were always tested against the error term provided by the variation among replicates (and, thus, not against significant interactions). This analysis provided information on the existence of differences among: 1) the grand means for the three pectin levels, 2) the grand means for the three flavoring levels, and 3) the grand means

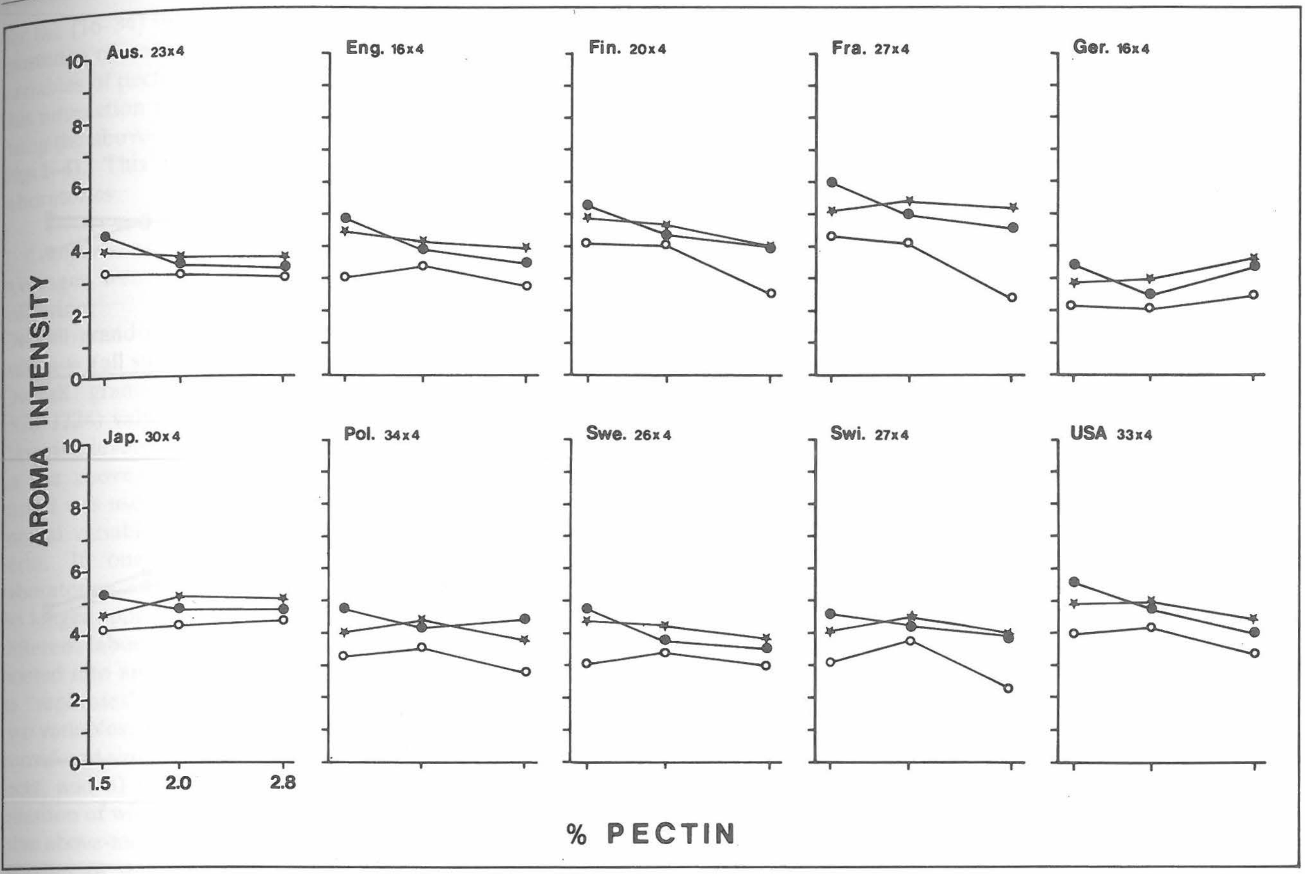

Fig. 1 *Pooled average aroma intensity vs. amount of pectin for each of three flavoring levels (o = "low", * = "medium", o = "high".* – Results of ten laboratories (Australia, England, Finland, France, Germany, Japan, Poland, Sweden, Switzerland, and USA).

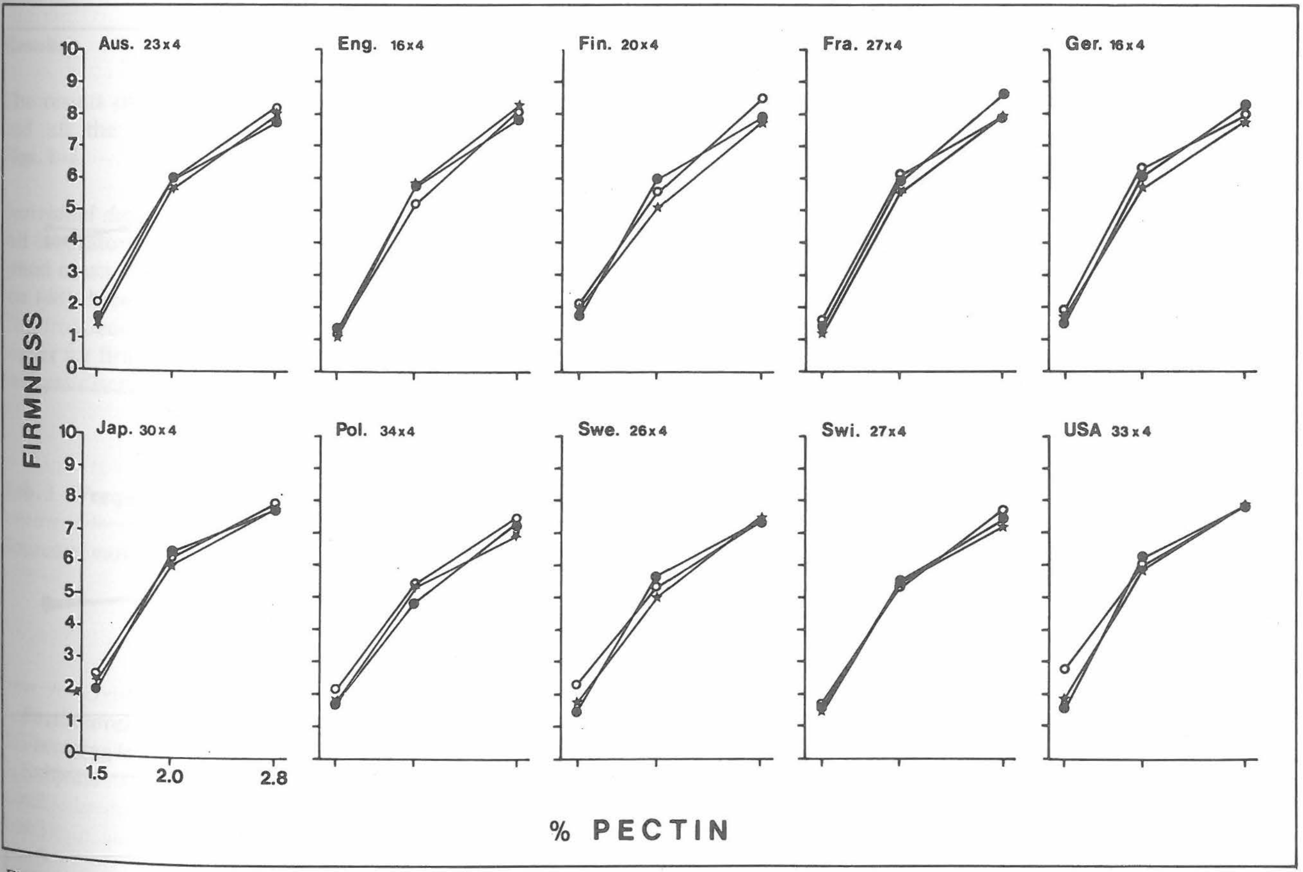

Fig. 2 *Pooled average firmness intensity vs. amount of pectin for each of three flavoring levels (o = "low" * = "medium" o = "high").* – Legend: see Figure 1.

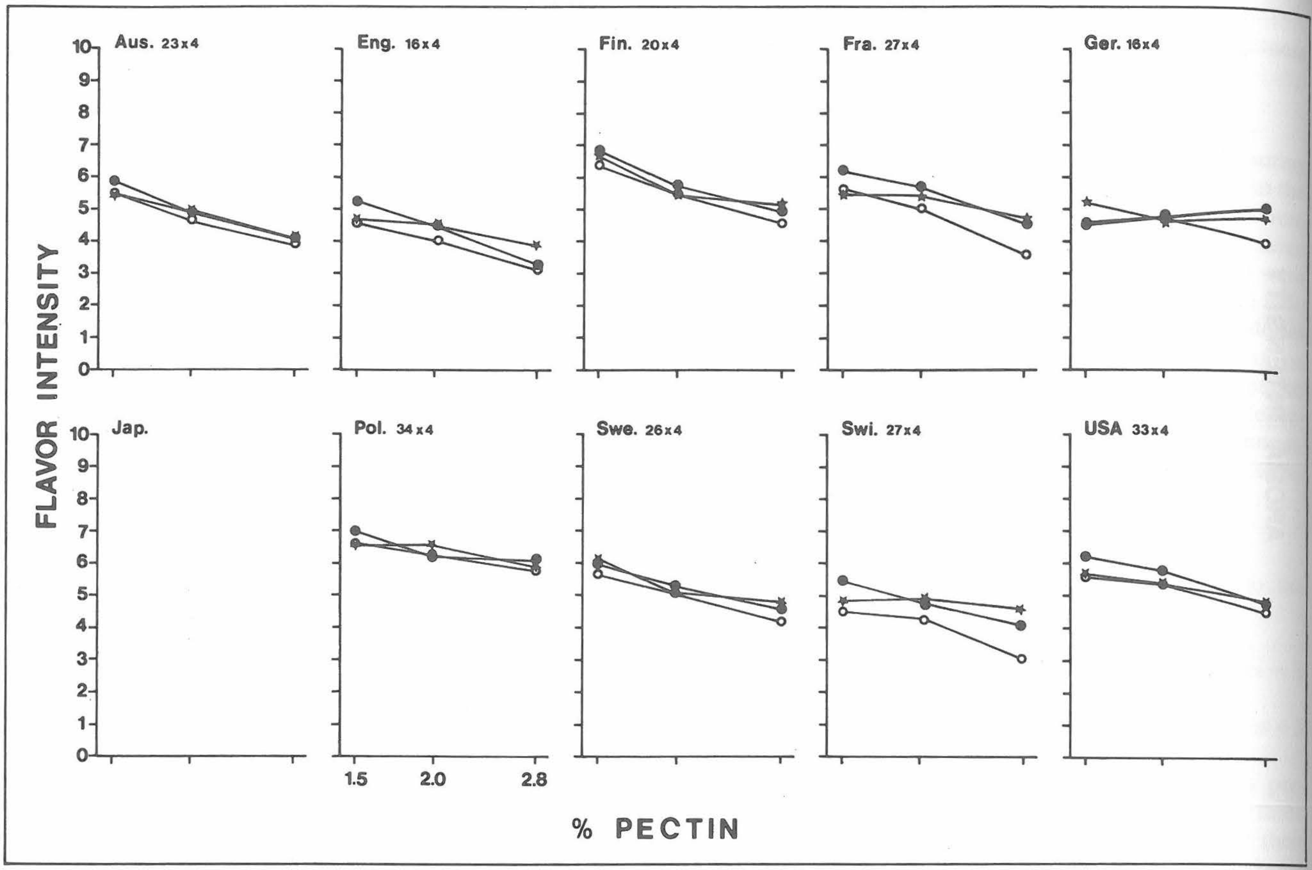

*Fig. 3 Pooled average flavor intensity vs. amount of pectin for each of three flavoring levels (o = "low", * = "medium", o = "high").* – Legend: see Figure 1.

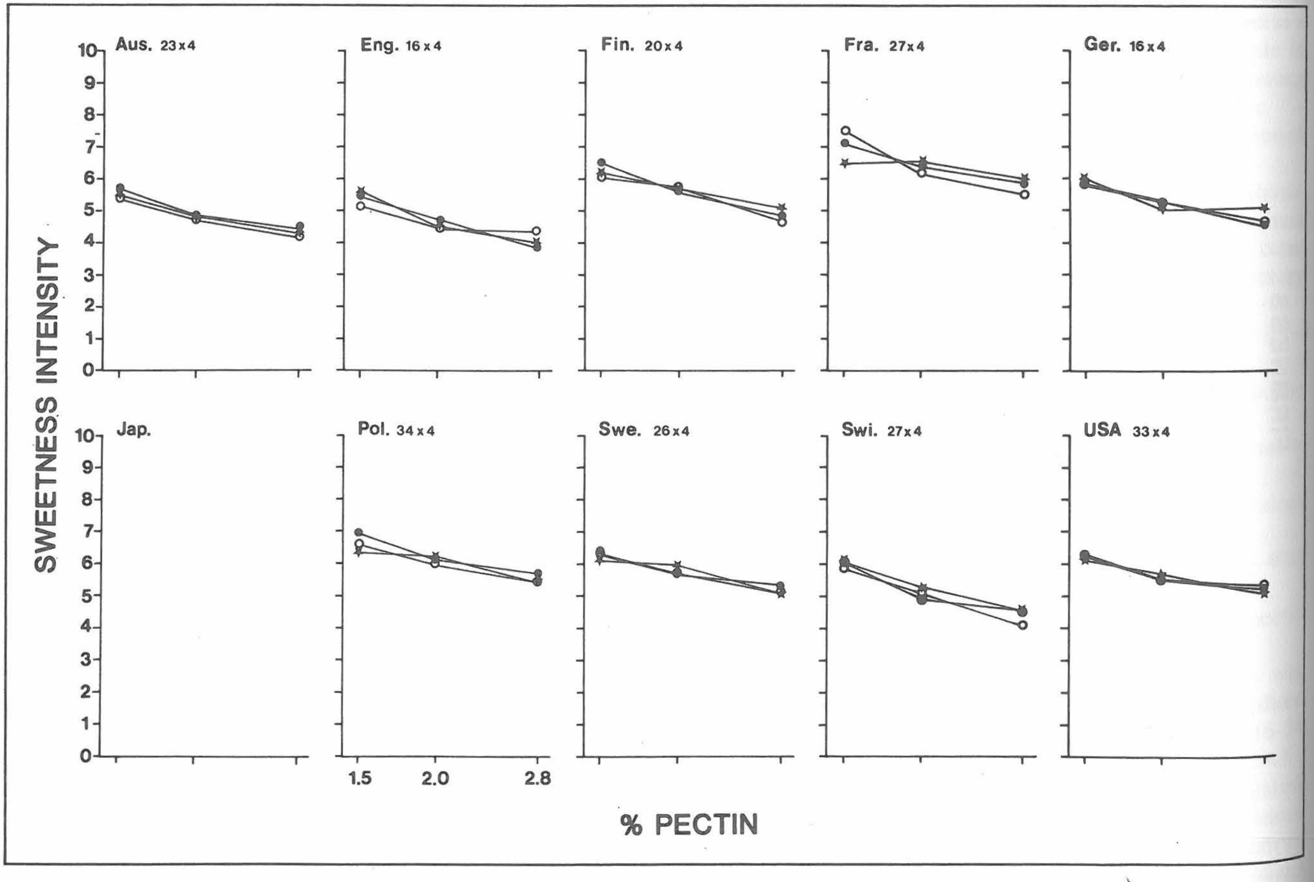

*Fig. 4 Pooled average sweetness intensity vs. amount of pectin for each of three flavoring levels (o = "low", * "medium", o = "high").* – Legend: see Figure 1.

for the (16–34) subjects. It also provided information on the existence of an interaction between, *e.g.*, the independent variables of pectin and flavoring content. The appearance of this interaction was seen from a "laboratory regression plot" using the above-mentioned pooled averages for samples (cf., **Figs 1–4**). This plot was compared to those from other laboratories.

3. Comparisons among laboratories

Averages were calculated and denoted according to the following:

Overall grand mean for a sample = average of (464–1008) values = (all subjects) × 4 reps.

Overall grand mean for a laboratory = average of (576–1224) values = (16–34 subjects) × 3 pectin levels × 3 flavoring levels × 4 reps.

In the above-mentioned balanced three-way ANOVA, a model was used, where subjects were considered an experimental variable, and replicates were used to give an error term. If one wants to include data from different laboratories, *i.e.*, different groups of subjects, this model is no longer applicable (since subject 1 is a different person at different laboratories, and, therefore, their data cannot be pooled into an average). Instead, subjects were considered as "replicates" in an unbalanced two-way ANOVA with the two variables: 1) samples (df = 8; *i.e.* all nine samples were considered simultaneously), 2) laboratories = groups of subjects, and 3) their interaction. The analysis answered the question of whether or not there were differences among the nine above-mentioned overall grand means for samples; and whether or not there were differences among the above-mentioned overall grand means for laboratories; and whether there was any interaction between sample and laboratory.

Results

The results of all ANOVAs are summarized in Tables 3–5, and all the "laboratory regression" plots are given in **Figs. 1–4**.

Individual data

All laboratories found highly significant differences among grand means for subjects as shown by the summary table of the (48) three-way ANOVAs (**Tab. 3**).

The frequencies of discriminators (**Tab. 4**) were very much higher for firmness than for the other sample attributes. The subjects discriminated most often between samples with different pectin contents, and more rarely between different flavoring contents. The relative frequencies of discriminators were similar at the different laboratories: All subjects discriminated firmness, and about half of them discriminated the other attributes.

Pooled data

Each laboratory's results (pooled data) regarding sample differences will be presented separately for each experiment. The conclusions of the three-way ANOVAs have been summarized in **Tab. 3**. The grand means corresponding to the main effects "pectin levels" and "flavoring levels" in these analyses are shown in Figures 5 and 6, respectively, whereas the interactions between pectin and flavoring content are given in **Fig. 1–4**.

The results of the unbalanced two-way ANOVAs, including the data of all laboratories simultaneously, are given in **Tab. 5**. For all attributes, highly significant differences were shown among the samples as well as among the laboratories, *i.e.* groups of subjects. The presence of significant differences among the overall grand means of the nine samples is in agreement with the conclusions reached for each laboratory's data (see below). The type and size of the differences among laboratories will be presented below.

Aroma intensity (Experiment A)

Pectin level (main effect): Nine of the ten laboratories showed highly significant differences among pectin levels (**Tab. 3**), whereas the tenth (Jap.) showed no significant difference which is easily confirmed by the upper left graph of **Fig. 5**. The size of the differences varied somewhat among the laboratories. Six of the nine gave the same rank order among the grand means: "hard < medium < soft", whereas two others (Pol. and Swi.) gave the order: "hard < soft < medium" with only a small difference between "soft" and "medium" which makes their results similar to those of the first six. Only one laboratory (Ger.) gave a rank order opposite to the other nine: "medium < soft < hard". No explanation can be given for this deviating result.

Flavoring level (main effect): All ten laboratories showed highly significant differences among flavoring levels, with practically identical magnitudes of the differences (**Fig. 6**). As can be seen from the upper left graph ("aroma") of **Fig. 6**, one of the ten (Pol.) gave the rank order "low < medium < high", whereas all others gave the order "low < medium = high".

Interaction: All laboratories except two (Eng. and Ger.) showed a significant interaction between the flavoring and pectin content. **Fig. 1** reveals that all ten show a cross-over

Tab. 3 Frequencies of laboratories obtaining significant differences in three-way ANOVAs

Source of variation	df	Exp. A		Exp. B		Exp. C				Exp. D			
		aroma intensity		firmness		flavor intensity I		sweetness		flavor intensity II		sourness	
		N.s.[1]	Sign.	N.s.	Sign.	N.s.	Sign.	N.s.	Sign.	N.s.	Sign.	N.s.	Sign.
1. Pectin level	2	1	9		10	1	8		9		5		5
2. Flavoring level	2		10	2	8	2	7	7	2		5	4	1
3. Subjects	15 – 33		10		10		9		9		5		5
1 × 2	4	2	8	2	8	5	4	8	1	4	1	3	2
1 × 3	30 – 66	9	1		10	2	7		9	1	4		5
2 × 3	30 – 66	9	1	10		9	1	9		3	2	5	
1 × 2 × 3	60 – 132	9	1	10		9	1	9		4	1	5	

1) N.s. and sign. denotes significance levels >5% and 5%, respectively.

Tab. 4 Frequencies of «discriminators»

Laboratory (No. subjects)	Source of variation [1] (Independent variable)	Frequency of discriminators[2] for attribute					
		aroma intensity	firmness	flavor intensity I	sweetness	flavor intensity II	sourness
1. Aus. (23)	1. pectin level	1	23	14	10	–[3]	–
	2. flavor level	2	2	1			
	1 × 2	1					
2. Eng. (16)	1. pectin level	1	16	6	9	–	–
	2. flavor level	3			1		
	1 × 2	2			1		
3. Fin. (20)	1. pectin level	6	20	12	12	10	12
	2. flavor level	5	2	1	1	3	1
	1 × 2		3	1	1	3	1
4. Fra. (27)	1. pectin level	6	27	10	10	11	18
	2. flavor level	11	2	4	1	2	
	1 × 2	3	1	1	1	4	
5. Ger. (16)	1. pectin level	2	16	4	5	7	6
	2. flavor level	5	1			2	
	1 × 2	1	2	2		1	1
6. Jap. (30)	1. pectin level	3	30	–	–	–	–
	2. flavor level	2	3				
	1 × 2	1	3				
7. Pol. (34)	1. pectin level	3	34	15	19	–	–
	2. flavor level	9	4	4	4		
	1 × 2		1	1	2		
8. Swe. (26)	1. pectin level	4	26	15	13	12	21
	2. flavor level	6	2	2		2	2
	1 × 2	1	4	1			2
9. Swi. (27)	1. pectin level	3	27	10	11	10	11
	2. flavor level	11		6	2	3	
	1 × 2	3		1		1	1
10. USA. (33)	1. pectin level	7	33	14	14	–	–
	2. flavor level	7	2	2			
	1 × 2	2	2	1	1		
Total	1. pectin level	36	252	100	103	50	68
	2. flavor level	61	18	20	9	12	3
	1 × 2	14	16	7	6	9	5
Total No. subjects participating		252	252	222	222	116	116

1) In two-way analyses of variance of individual data.
2) Discriminator = subject for whom a significant difference was found ($p < 5\%$) for the independent variable 1 or 2. Cf. text.
3) – indicates that these experiments were not made by this laboratory.

interaction: A lower average aroma intensity of two of the samples containing the highest flavoring level compared to those with the middle concentration. In addition, data from two laboratories (Fin. and Fra.) suggest a magnitude interaction consisting of a higher slope of the data for the lowest flavoring concentration.

For the aroma intensity data, where the differences between the ten laboratories' data were the largest, it would be useful to evaluate the size of variation among the plots. One way to do so would be to compare this variation with a corresponding within-laboratory variability. Since the experiments were done only once, no within-laboratory repeatability measure with the same number of subjects can be calculated. Instead, the following simulation of a possible group-to-group variation was made. The USA laboratory was randomly selected from the two laboratories with the highest number of subjects. The 33 subjects were divided into two sub-groups: one consisting of 16 subjects randomly selected out of the 33, and another of the remaining 17 (= complementary group). This randomisation was repeated

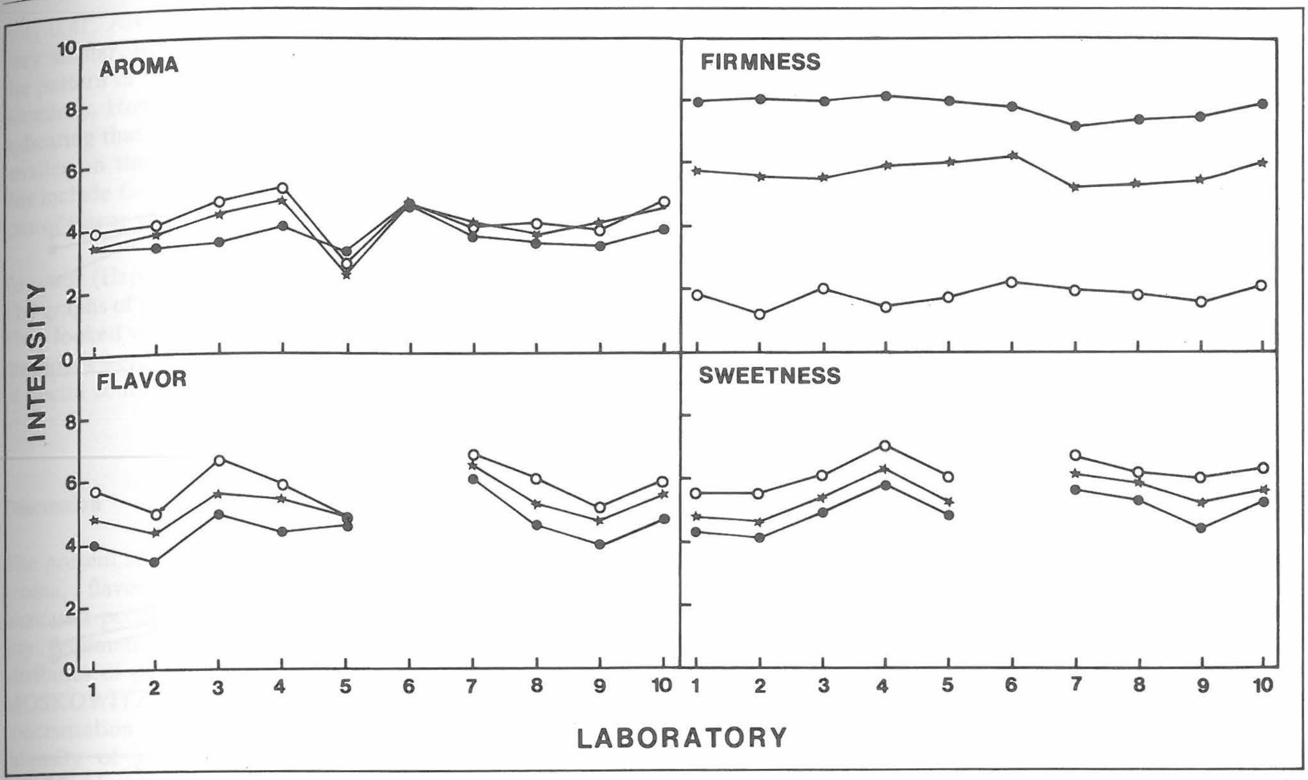

Fig. 5 *Grand means of each pectin level (o = "soft", * = "medium", o = "hard").*
Laboratories:
1 = Aus. (n = 276)
2 = Eng. (n = 192)
3 = Fin. (n = 240)
4 = Fra. (n = 324)
5 = Ger. (n = 192)
6 = Jap. (n = 360)
7 = Pol. (n = 408)
8 = Swe. (n = 312)
9 = Swi. (n = 324)
10 = USA (= 396)

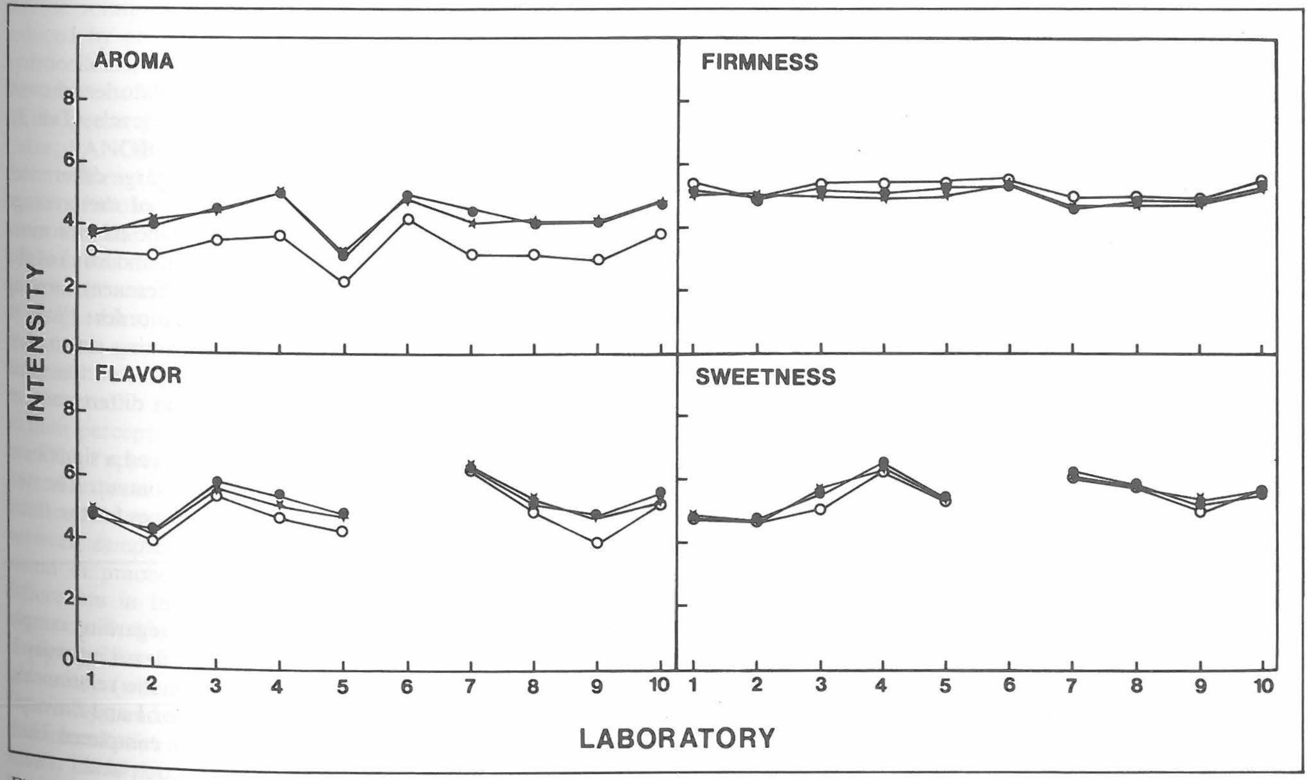

Fig. 6 *Grand means of each flavoring level (o = "low", * = "medium", o = "high").* – Legend: see Figure 5.

Tab. 5 Comparisons of different laboratoires
Results of unbalanced two-way analyses of variance

Sensory attribute (No. laboratories)	Source of variation					
	Samples (S)		Labs. (L)		S×L	
	F	sign. level[1)]	F	sign. level	F	sign. level
Aroma intensity (10)	39.4	***	28.4	***	1.21	*
Firmness (10)	1503	***	11.2	***	2.00	***
Flavor intensity I (9)	36.0	***	45.1	***	<1	n.s.
Sweetness (9)	23.3	***	27.2	***	<1	n.s.
Flavor intensity II (5)	19.2	***	13.4	***	<1	n.s.
Sourness (5)	32.4	***	8.6	***	<1	n.s.

1) n.s., *, **, *** denotes significance levels >5%, 5%, 1%, and 0,1%, respectively.

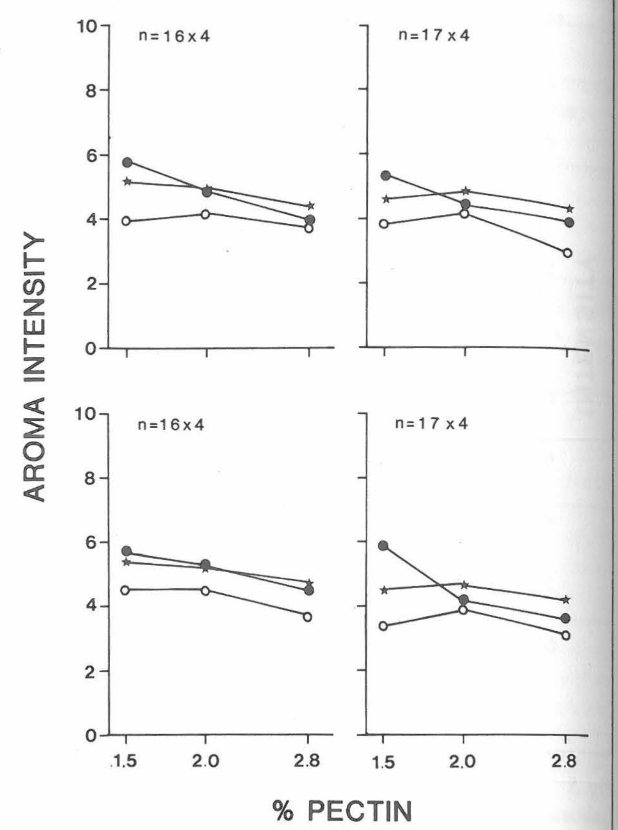

Fig. 7 Examples of random group-to-group variation within a laboratory (USA)
Upper graphs: Randomisation 1= most similar sub-groups
Lower graphs: Randomisation 2= least similar sub-groups.

four times, and the extreme outcomes are shown in **Fig. 7**. Since the subjects were randomly selected from a larger, supposedly homogeneous, group, the procedure was considered to give a measure of a possible within-laboratory variance, including variation in interaction. The differences among the ten plots in **Fig. 1** appear to be within the range of this within-laboratory variation.

Firmness (Experiment B)
Pectin level (main effect): As anticipated, all ten laboratories showed highly significant differences among pectin levels, with the same expected rank order: "soft < medium < hard" (**Fig. 5**). The magnitudes of the differences were very similar.
Flavoring level (main effect): Eight of the ten laboratories showed significant differences among flavoring levels (**Tab. 3**) with practically identical magnitudes of the differences (**Fig. 6**). All eight gave the rank order "high ≃ medium < low" (upper right graph of **Fig. 6**). Of the remaining two, (Swi.) also gave this rank order although the differences were not significant. The tenth laboratory (Eng.) showed practically identical grand means for the three levels.
Interaction: All laboratories except Pol. and Swi. showed a significant interaction between the flavoring and pectin content. **Fig. 2** shows that the type of interaction was different among the eight, from intensity interaction for USA to different cross-over interactions for all the other seven.

Flavor intensity I (Experiment C)
Pectin level (main effect). With one exception (Ger.), the conclusions regarding the effect of pectin level were the same as for sweetness below (**Fig. 5**). In contrast to the other eight (Jap. did not perform these evaluations), this laboratory showed no significant differences. No explanation can be given for this deviating result.
Flavoring level (main effect): For all nine laboratories, the conclusions regarding the effect of the flavoring level were practically the same as for aroma, only the differences were smaller (the lower left graph of **Fig. 6**).
Interaction: Three of the four significant interactions (Fra., Ger., Pol., and Swi.) look similar to those for the aroma intensity (**Fig. 3**).

Sweetness (Experiment C)
Pectin level (main effect): All nine laboratories showed highly significant differences among pectin levels (**Tab. 3**), giving the rank order "hard < medium < soft".
The lower right graph of **Fig. 5** shows fairly large differences among the grand means. Also, the sizes of the average differences are remarkably similar for all nine laboratories.
Flavoring level (main effect): Only two (Pol. and Swi.) of the nine laboratories showed significant differences among flavoring levels. Both gave the same rank order: "low < medium = high". Also, one of the remaining seven that showed no significant differences (Fin.), indicated the same rank order, whereas all the others gave no differences, as seen from the lower right graph of **Fig. 6**.
Interaction: None of the nine data sets showed a significant interaction between flavoring and pectin content. The ten graphs of **Fig. 4**, all show almost parallel curves for the three flavoring levels.

Flavor intensity II (Experiment D)
With one exception (Ger.), the conclusions regarding sample differences were practically the same as for flavor intensity I. Linear regression "curves" were fitted for the relationship between the nine grand means of the flavor I and flavor II evaluations for the five laboratories that completed both experiments. Four of the five gave high correlation coefficients (≥0.92), whereas the fifth (Ger.) had a coefficient of

only 0.41. Also, the four graphs (not shown here), looked very similar, implying that the same conclusions regarding the pattern of the sample differences were obtained on both occasions. However, the slopes of the fitted lines were <1, indicating that the sample differences were perceived to be smaller on the second occasion. Possible explanations for this include fatigue of the subjects, or sample changes, *e.g.*, loss of flavor with time.

Sourness (Experiment D)
The graphs of the sourness intensity data are not shown here. They looked very similar to these of the sweetness data, only the slopes were steeper, indicating a more pronounced effect of pectin content on sourness.

Discussion

The present study showed a significant decrease in average aroma, flavor, sweetness and sourness intensity with increased pectin content. No reference has been found to any systematic study of sensory aroma, flavor, or taste attributes of pectin gels of different pectin concentrations. MOSKOWITZ and ARABIE (8) found that an increased concentration of carboxymethylcellulose decreased the taste intensity of glucose (sweetness), citric acid (sourness), sodium chloride (saltiness), and quinine sulphate (bitterness) in water solutions. Similarly, CHRISTENSEN (9) showed that carboxymethylcellulose decreased saltiness of sodium chloride in water. PANGBORN *et al.* (10), PANGBORN and SZCZESNIAK (11), and PANGBORN *et al.* (12) studied the effect of five hydrocolloids (hydroxypropylcellulose; carboxymethylcellulose-low-viscosity; carboxymethylcellulose-medium-viscosity; sodium alginate; and xanthan) on aroma, flavor, and taste intensity of different odor, flavor, and taste compounds. With some interesting exceptions for saccharin, an increased concentration of the thickener in all cases caused a decrease in aroma, flavor, and taste intensity. PANGBORN *et al.* (10) showed that sweetness of sucrose solutions and sourness of citric acid solutions was reduced by an increased concentration of the mentioned hydrocolloids, and that the effect was more pronounced for the sourness. The present results on sweetness and sourness of the pectin gels are in agreement with this observation. Later, PANGBORN *et al.* (12) showed the same effects of the mentioned hydrocolloids in an orange drink. The present results of aroma and taste of a semi-solid food are in agreement with the mentioned findings.

The subjects in the present study were not trained to use the scale in the same way by, *e.g.*, presenting reference samples of assigned intensities or intensity differences, or by reaching consensus among panel members about the scale usage. Instead, the subjects were free to use any part of the scale they wanted to express intensities (as long as it corresponded to their perceptions, of course). As expected, some subjects used the "upper" part of the scale, and others the "lower" part to express intensity *level* of a sample. They also used different distances of the scale to express intensity *differences* between samples. Such variations in scale usage, experienced in practically all sensory experiments, are due to differences in individual behavior and/or sensitivity (see, *e.g.*, refs. 4, 13). The differences in scale usage very likely imply significant differences between subjects *within* each laboratory (when the data of each laboratory are treated separately), and *between* laboratories (when the data of several laboratories are included in one analysis). In addition, it is very likely that interactions including subjects will also be significant.

The present study gave highly significant subject differences in all within-laboratory ANOVAs. They are considered due to the above discussed differences in scale usage. It also gave highly significant differences between laboratories. Until shown otherwise, it is considered that these differences are also due to differences in scale usage. To see whether some "genuine" differences between laboratories possibly remained that could not be "explained" by differences in scale usage, a reduction of data from "metric" to rank order level was done. Using the regression plots and tables, the general pattern, i.e. rank order and trends in the plots, of a laboratory's average response was compared visually with those of the others.

With few exceptions (*cf.* results) all ten laboratories gave the same aroma intensity pattern for the differences among the nine samples. The differences in the appearance of the ten plots (**Fig. 1**) was greater for this attribute than for the others, most likely a result of the relatively lower general intensity of the aroma of the product. Compared to the firmness differences, evident to all subjects, and compared to the very strong sweetness of the product, the aroma differences were faint and difficult to detect. This statement is based on comments given by experimenters and many subjects; it cannot be "proven" from the data presented here. On the contrary: since the aroma and sweetness plots have about the same position near the centre of the intensity axis, aroma and sweetness intensities appear to be similar. This is, however, highly misleading as the sweetness was very high, and the aroma much weaker. The position of the plots near the centre of the scale is probably due to the well-known "centre effect" which means that subjects have a tendency to use the whole scale to "measure" sample differences rather than to use it as an absolute scale (see, *e.g.*, refs. 14, 15).

For all remaining attributes, the ten laboratories gave almost identical response patterns of the sample differences across pectin and flavoring contents. The interactions looked different for the different laboratories, however; but generally they were small.

The high degree of similarity of the response patterns supports the conclusion that there were no between-laboratory differences other than the "normal subject-to-subject" variation encountered in every sensory experiment of this kind. Rather, the high degree of similarity of both patterns and magnitudes of the differences should be pointed out. It indicates that, with a large number of evaluations (subjects and/or replicates), the differences in average intensities approach "true" values which are independent of the laboratory.

The aroma and flavor evaluations were made separately; experiment A and C, respectively. Generally, the flavor plots showed smaller differences between flavoring levels than the aroma plots, possibly indicating that flavor = "aroma-by-mouth" is more difficult to discriminate than "aroma-by-nose". Most of the plots look like an "average" or "composite" of the aroma and sweetness plots, with some being closer to the aroma data, *e.g.*, Swi., and others to the sweetness data, *e.g.*, Aus.

It is difficult to define or explain "total flavor" to subjects; should it be "aroma-by-mouth" disregarding the basic tastes, or a "composite" of aroma and basic tastes? Supposedly, in English and French, with the three separate words aroma, flavor and taste, it should be easier to distinguish the meaning of "total flavor intensity" from aroma and basic tastes than in other languages which lack the word flavor, *e.g.* Ger. and Swe. Then, the flavor data from the English and French speaking laboratories Aus., Eng., Fra., Swi., and USA would all have flavor data more similar to the aroma data, if flavor were equal to "aroma-by-mouth", or closer to the sweetness data if it were equal to aroma and sweetness together (weak aroma and high sweetness). However, neither of these patterns was seen in all five cases. Of the remaining four laboratories, one gave a plot that was similar to neither aroma nor sweetness data, possibly indicating that flavor was not defined to the subjects. This is supported by the fact that this laboratory obtained flavor intensity II data that were similar to data of the other laboratories.

The present data clearly show the need for a more stringent definition of a term like "total flavor intensity" in future studies. An interesting and extensive discussion of the meaning of the word flavor has recently been given by Rozin, 1982, confirming the above-mentioned difficulties.

Considering the large number of factors that may influence the outcome of a sensory experiment, the high degree of similarity among the results of the different laboratories reported herein is encouraging for the field of sensory analysis. It shows that there is a high likelihood that a carefully controlled sensory experiment performed by laboratories in different locations will lead to the same conclusions, despite, *e.g.*, language differences. The gel product used in the present study was very homogeneous, and the sensory "profile" simple; the philosophy behind this choice being: "if we cannot get agreement among results for such simple food, how can we get it for a more complex one"? Now that a high degree of agreement between results has been shown, the next step would be to involve a more complex stimulus.

Conclusions

Laboratories in ten countries with different languages and food habits performed the same, carefully controlled sensory experiment involving odor, taste and texture evaluations of pectin gels. Among the definite conclusions were: 1) there were no other differences between laboratories than the normal subject-to-subject variation encountered within-laboratory, 2) a high degree of similarity among the response patterns from the different laboratories, and 3) all laboratories established that the aroma, flavor and taste (sweetness and sourness) intensity decreased with increased pectin content. The magnitudes of the average differences (across pectin or flavoring contents) were remarkably similar among the laboratories. The high number of evaluations (hundreds) by each laboratory implies that a laboratory's data are good estimates of the "true" sample intensity differences from that laboratory. Considering the small variation among the estimates from the ten laboratories, they all seem to come from the same population.

Acknowledgements

We are thankful to Nestlé Products Technical Assitance Co. Ltd., La Tour-de-Peilz, Switzerland, for manufacturing and shipping a very large number of samples. Students of the College of Science and Engineering, Chuo University, Tokyo, Japan, are thanked for performing computer analyses of data. We also thank all subjects who so patiently put up with our long-lasting experiments, and the following persons for skilful help with the work: *D.J. Best,* CSIRO, Australia; *Cynthia Thurston,* FIR, Norwich, England; *Saija Kurunmäki,* University of Helsinki, Finland; *M. Chapon,* ENSBANA, Dijon, France; *Y. Andersson, J. Andersson, G. Hall, B. Karlström, B. Olofsson* and *Gerd Snickert,* SIK, Göteborg, Sweden; *Mrs Voirol,* Nestlé, Vevey, Switzerland; *Lori Ringo,* University of California, Davis, USA.

References

1 JOHANSSON, B., DRAKE, B., PANGBORN, R.M., BARYLKO-PIKIELNA, N., and KÖSTER, E.P., Difference Taste Thresholds for Sodium Chloride Among Young Adults: An Interlaboratory Study, J. Food Science, *38,* 524–527 (1973)

2 LUNDGREN, B., PANGBORN, R.M., BARYLKO-PIKIELNA, N. and DAGET, N., Difference Taste Thresholds for Sucrose in Water and in Orange Juice: An Interlaboratory Study, Chemical Senses and Flavor, *2,* 157–176 (1976)

3 LUNDGREN, B., JONSSON, B., PANGBORN, R.M., SONTAG, A., BARYLKO-PIKIELNA, N., PIETRZAK, E., dos SANTOS GARRUTI, R., MORAES, M.A.C. and YOSHIDA, M., Taste Discrimination vs Hedonic Response to Sucrose in Coffee Beverage. An Interlaboratory Study. Chemical Senses and Flavor, *3,* 249–265 (1978)

4 DRAVNIEKS, A., McDANIEL, H.C., and POWERS, J.J., Comparison by Twelve Laboratories of the Odor Qualities of Nine Chemicals Sniffed from the Bottle and as Gas-Liquid Chromatography Effluents, J. Agric. Food Chem., *27,* 336–346 (1979)

5 POWERS, J.J., WARREN, C.B. and MASURAT, T., Collaborative Trials Involving Three Methods of Normalizing Magnitude Estimations, Lebensm.-Wiss. u.-Technol., *14,* 86–93 (1981)

6 DRANSFIELD, E., RHODES, D.N., NUTE, G.R., ROBERTS, T.A., BOCCARD, R., TOURAILLE, C., BUCHTER, L., HOOD, D.E., JOSEPH, R.L., SCHON, I., CASTEELS, M., COSENTINE, E., and TINBERGEN B.J., Eating Quality of European Beef Assessed at Five Research Institutes Meat Science, *6,* 163–184 (1982)

7 DAGET, N. and COLLYER, S., Comparison between quantitative descriptive analysis and physical measurements of gel systems and evaluation of the sensorial method, J. Texture Studies *15,* 227–245 (1984)

8 MOSKOWITZ, H.R. and ARABIE, P., Taste Intensity as a Function of Stimulus Concentration and Solvent Viscosity, Journal of Texture Studies, *1,* 502–510 (1970)

9 CHRISTENSEN, C.M., Texture-taste interactions, Cereal Foods World *22,* 243–244, 256 (1977)

10 PANGBORN, R.M., TRABUE, I.M. and SZCZESNIAK, A., Effect of Hydrocolloids on Oral Viscosity and Basic Taste Intensities, Journal of Texture Studies, *4,* 224–241 (1973)

11 PANGBORN, R.M. and SZCZESNIAK, A., Effect of Hydrocolloids and Viscosity on Flavor and Odor Intensities of Aromatic Flavor Compounds, Journal of Texture Studies, *4,* 467–482 (1974)

12 PANGBORN, R., GIBBS, Z.M. and TASSAN, C., Effect of Hydrocolloids on Apparent Viscosity and Sensory Properties of Selected Beverages, Journal of Texture Studies, *9,* 415–436 (1978)

13 PANGBORN, R.M., Individuality in Response to Sensory Stimuli. In: Criteria of Food Acceptance - How Man Chooses What He Eats. A symposium. J. Solms and R.L. Hall (eds.), Forster Publ. Ltd., Zurich, pp. 177–219 (1981)

14 POULTON, E.C., Quantitative/Subjective Assessments Are Almost Always Biased, Sometimes Completely Misleading, Br. J. Psychology, *68,* 409–425 (1977)

15 McBRIDE, R.L., Range Bias in Sensory Evaluations., J. Food Technology, *17,* 405–410 (1982)

16 ROZIN, P., "Taste-smell confusions" and the duality of the olfactory sense, Perception & Psychophysics, *31,* 397–401 (1982)