

Supplementary file

The normalized sensory attributes of the various experts are visualized by heatmap in Figure S.1. The whole figure consists of 6*6 subplots. Each subplot represents the sensory attributes of a specific yogurt sample. There are 22 columns, 10 rows in a heatmap, representing 22 sensory attributes values evaluated by 10 experts of a sample. Visualization by heatmap is helpful for explicit characteristics display of a sample.

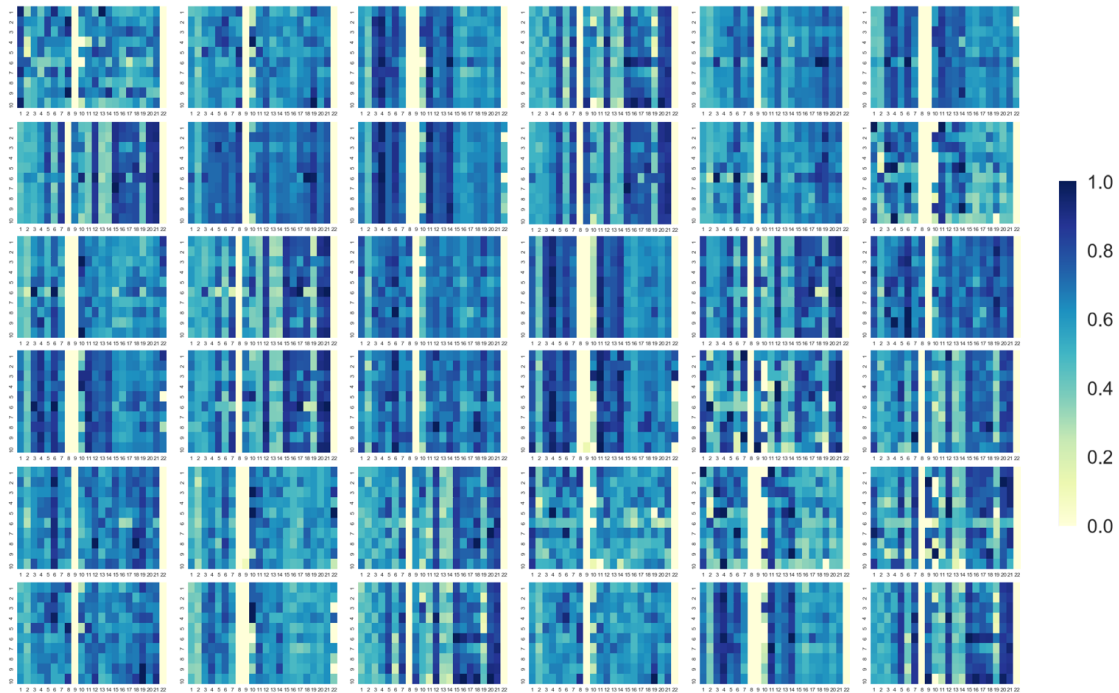


Figure S.1. Detailed sensory evaluation results of 36 yogurt samples assessed by 10 experts on 22 sensory attributes after normalization

Further expert performance analysis and data quality validation is discussed in the following paragraphs.

A scatter diagram composed of p-values and MSE-values is plotted in Figure S.2. These values are derived from the scoring results of each assessor using one-way ANOVA analysis. MSE-values, representing the repeatability ability, are plotted along the x-axis, and p-values, representing the discrimination ability, are plotted along the y-axis. Each point in the plot represents the evaluation of an assessor for a specific attribute. The closer a point is to the origo (lower values for both p and MSE), the better an assessor has performed. It is apparent from this plot that most assessors have good discrimination ability and repeatability ability.

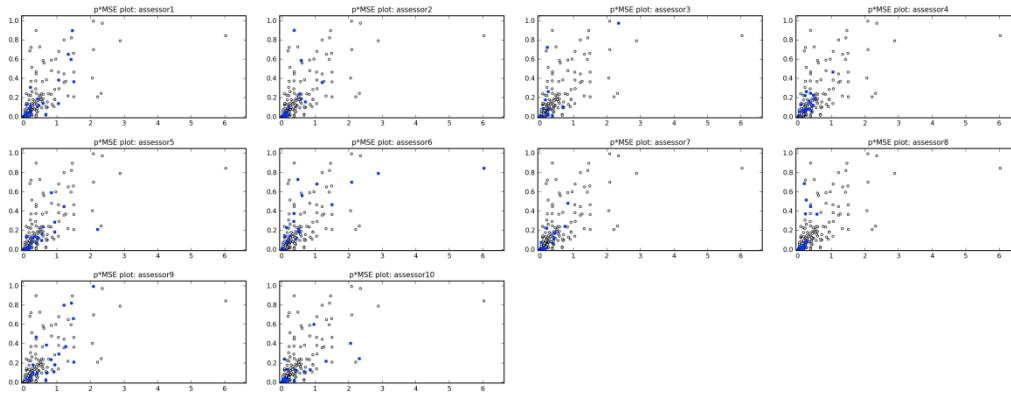


Figure S.2. P*MSE plot of the experts as a blind reference in three ratings

F-value for each assessor, calculated by one-way ANOVA is shown in Figure S.3. Each vertical line represents the F-value of an assessor on a specific attribute evaluation. High F-value means the assessor has a strong discrimination ability.

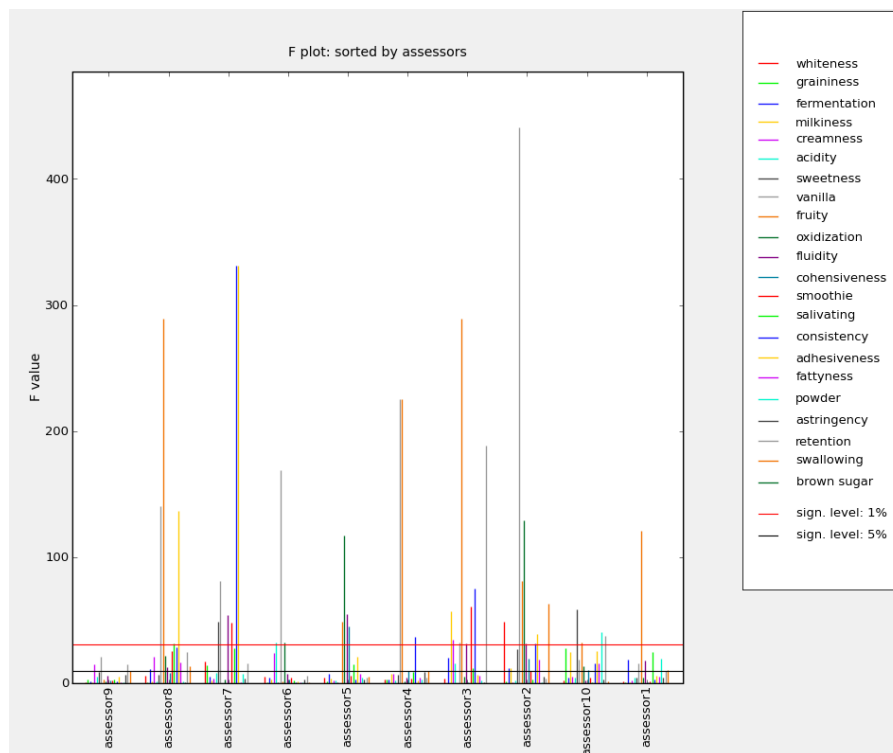


Figure S.3. F plot of the experts as a blind reference in three ratings

The results of the correlation analysis are presented in Figure S.4., which provides an overview of assessors' agreement of each attribute. The plot indicates that most evaluators reach agreement on all 22 attributes in the training session.

Tucker-1 plots are widely used to check whether assessors agree with each other on the scoring of attributes. The Tucker-1 plot of training session and of whole 36 samples is shown in Figure S.4.

Points accumulated on the outer circle in Figure S.4(a) indicates that assessors reach agreement on all 22 attributes in the training session. Points majorly concentrated on the same side and distributed between 50% and 100% explained variances in Figure S.4(b) indicates that the assessors reach agreement on all attributes of 36 samples.

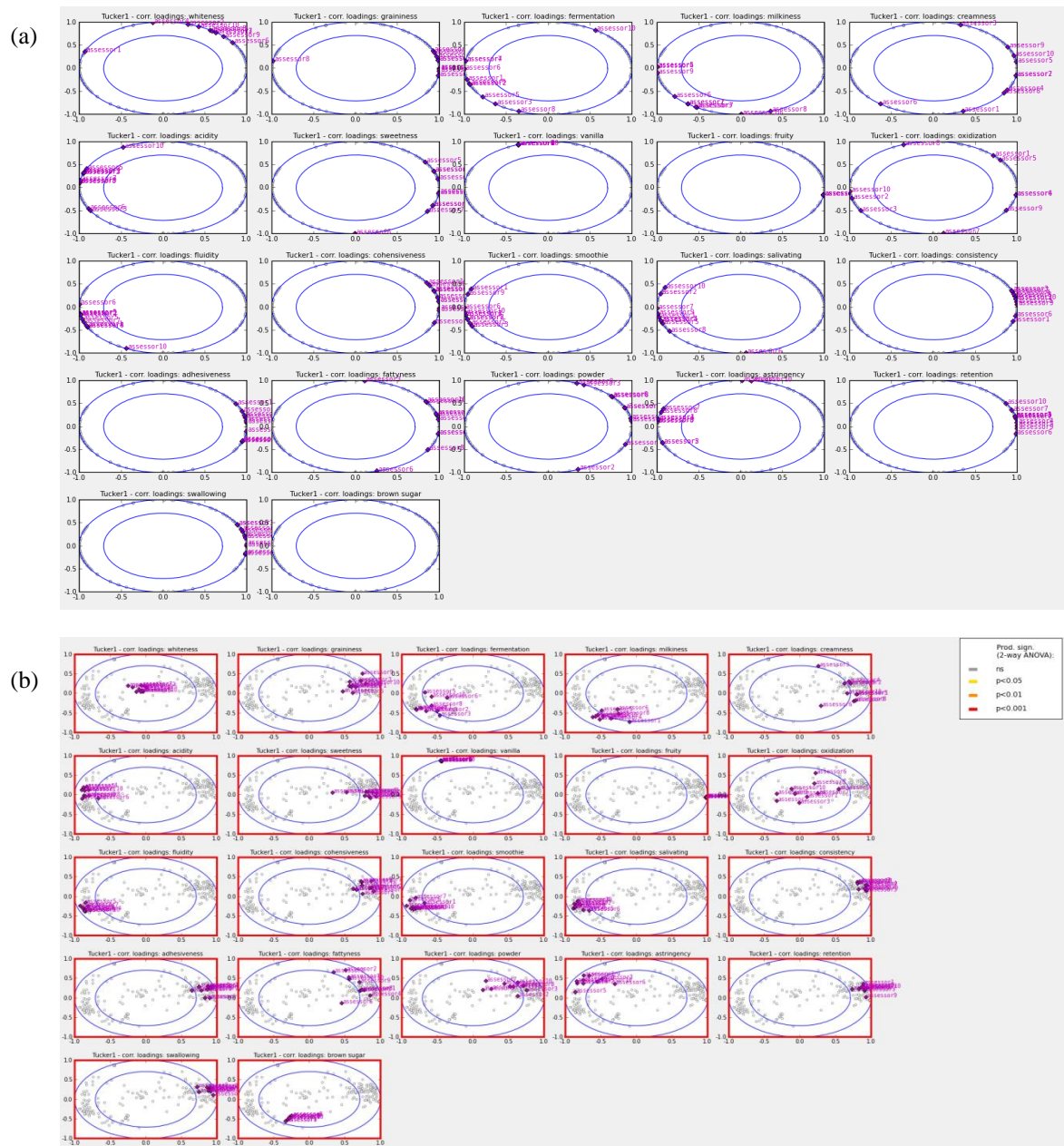


Figure S.4. Tucker-1 plot for 22 attributes (a) of blind references (b) of 36 samples

A two-way ANOVA provides a quick overview on the overall panel performance, which is shown in Figure S.5. From the bar charts, we can see that both samples and assessors have significant ($P < 0.05$) influence on 22 attributes.

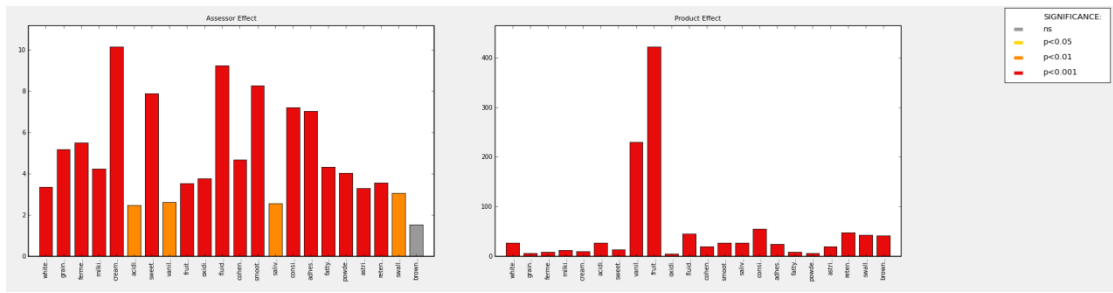


Figure S.5. 2-way ANOVA analysis of 36 samples