

Estimating affective taste experience using combined implicit behavioral and neurophysiological measures

A.-M. Brouwer, T. J. van den Broek, M. A. Hogervorst, D. Kaneko, A. Toet, *Member, IEEE*, V. Kallen, J. B. F. van Erp

Abstract—We trained a model to distinguish an extreme high arousal, unpleasant drink from regular drinks based on a range of implicit behavioral and physiological responses to naturalistic tasting. The trained model predicted arousal ratings of regular drinks, highlighting the possibility to estimate affective experience without having to rely on subjective ratings.

Index Terms—Affect sensing and analytics, Customer experience measurement, Nonverbal synthesis, Physiological measures, Tasting

1 INTRODUCTION

Knowledge of individuals' affective experience associated with eating or drinking certain foods or drinks is of great interest to the food industry as well as other parties involved in developing food products [1] and those promoting healthy eating behaviour. As reviewed in [2] and [3], affective food experience is usually measured using verbal questionnaires. While this is a relatively cheap method with high face validity, it suffers from certain drawbacks. Verbal descriptions are discontinuous, prone to demand characteristics, may suffer from response biases, and may not cover subconscious processes [2], [3], [4]. Also, the 'emotional lexicon' strongly varies between individuals [1], [5] and slight variations in phrasing can have (unintended) effects on the results [6]. Finally, repeatedly asking individuals about their food experience will interfere with the food experience itself, different from what is experienced in a natural, daily life situation. Measuring unconscious (implicit) responses in addition to self-reports may help us to better understand consumers' food-evoked

emotions and possibly predict their future food choice behavior [7], [8].

Physiological, implicit measures have been studied before in relation to food experience elicited by tasting [2] [3]. High heart rate and skin conductance or electrodermal activity (EDA) have been reported for unpleasant tastes [9], [10], [11] though (trends of) the opposite have been found as well [12]. From a physiological point of view, variables reflecting activity of the autonomic nervous system such as EDA, heart rate and also pupil diameter and heart rate variability are expected to be more closely associated with arousal rather than pleasantness (valence) [13]. The reported relations between valence and these variables may have been mediated by arousal, where valence and arousal may have been associated in the stimuli used. While existing tasting studies often omit the arousal axis of affect, [12] included ratings of arousal. They indeed found the expected positive relation between EDA and arousal ratings.

A variable that could be expected to convey valence is EEG (electroencephalography) frontal alpha asymmetry. This measure has been shown to be related to approach-avoidance motivation [14], which in the context of food, may be considered as a proxy for valence. Expected alpha asymmetry effects have been found in viewing and cooking food [15],[16] but, as far as we are aware, not in tasting.

We recently conducted a study with the aim of comparing these measures all in the same context of tasting drinks from cups [17]. We designed our experiment to mimic natural tasting as well as possible. This means that rather than employing strict laboratory conditions and rigid sampling regimes (e.g. [10],[11],[12]), participants tasted common drinks, they knew what drink they were about to taste, and apart from being instructed to take one sip at a designated time point, and to take sips in a similar way, they were not instructed in other ways (e.g. on exactly how long to keep the sip in the mouth). With this approach, we cannot disentangle effects of different sensory processes (vision, olfaction and taste) and expectation, and it will introduce

- A.-M. Brouwer is with The Netherlands Organization for Applied Scientific Research TNO, Kampweg 55, 3769 DE Soesterberg, The Netherlands. E-mail: anne-marie.brouwer@tno.nl.
- T.J. van den Broek is with The Netherlands Organization for Applied Scientific Research TNO, Utrechtseweg 48, 3704 HE Zeist, The Netherlands. E-mail: tim.vandenbroek@tno.nl.
- M. A. Hogervorst is with The Netherlands Organization for Applied Scientific Research TNO, Kampweg 55, 3769 DE Soesterberg, The Netherlands. E-mail: maarten.hogervorst@tno.nl.
- Daisuke Kaneko is with Kikkoman Europe R&D Laboratory B.V., Nieuwe Kanaal 7, 6709 PA Wageningen, The Netherlands. E-mail: d.kaneko@kikkoman.nl.
- Alexander Toet is with The Netherlands Organization for Applied Scientific Research TNO, Kampweg 55, 3769 DE Soesterberg, The Netherlands. E-mail: lex.toet@tno.nl.
- V. Kallen is with The Netherlands Organization for Applied Scientific Research TNO, Kampweg 55, 3769 DE Soesterberg, The Netherlands. E-mail: victor.kallen@tno.nl.
- J.B.F. van Erp is with The Netherlands Organization for Applied Scientific Research TNO, Kampweg 55, 3769 DE Soesterberg, The Netherlands and Human Media Interaction, University of Twente, Enschede, The Netherlands. E-mail: jan.vanerp@tno.nl.

more noise due to (slightly varying) body movements. On the other hand, it is more compatible with a real-life tasting situation. Also, this naturalistic approach gained us the potentially interesting, implicit variable of sip size. In the remainder, with ‘tasting’, we refer to the complete set of sensory and anticipatory effects that go with tasting a known drink. We refer to Verastegui-Tena and colleagues for physiological studies that specifically look at the role of expectation in tasting [8]. In our study, participants were asked to taste eight regular drinks as well as a vinegar solution, that was a priori expected to be strongly disliked (high arousal and low valence compared to the regular drinks). We recorded the physiological implicit measures as discussed above, the implicit behavioural measure of sip size, and ratings of valence and arousal. These were all tested for their sensitivity to distinguish between the regular drinks and the vinegar solution. Physiological measures were also examined with respect to their development over time. EDA, heart rate and pupil diameter showed consistent temporal patterns: increasing after presentation of the name of the drink which was the sign to take one sip, peaking between 5-15 sec and then decreasing again. The discriminative power of distinguishing between the vinegar solution and the regular drinks as expressed by a z-score was highest for sip size ($z=19.5$), followed by rated valence ($z=14.8$), rated arousal ($z=13.4$), heart rate ($z=11.0$), EDA ($z=8.0$), pupil diameter ($z=4.3$) and EEG frontal asymmetry ($z=2.6$). There was no significant difference between vinegar solution and regular drinks for heart rate variability. For the regular drinks, positive correlations were found between arousal ratings and heart rate, and between rated arousal and heart rate variability (where the heart rate variability recorded during tasting the vinegar solution was at approximately the same level as the drink next highest in rated arousal, suggesting that a floor level may have been reached that prevented this variable from successfully distinguishing between the vinegar solution and the regular drinks). For regular drinks, there was also a positive relation between valence and heart rate, but further data inspection and analyses indicated that this was probably mediated by a positive relation between rated arousal and valence in regular drinks.

These results were encouraging with respect to the potential of implicit measures to estimate affective food experience. In the current study, we used the data of this same study to take a more quantitative approach and explore the potential of combining implicit measurements, covering the behavioral, physiological and neurophysiological domain, to estimate affective food experience.

A challenge in using implicit measures to estimate affective food experience, as well as in affective computing in general, is the lack of a ground truth measure of affect [18]. While we know that implicit measures such as the ones mentioned above are associated with affect, their relation is not clear-cut and context dependent. Therefore, when modelling affect using (combinations of) implicit measures, we need ground truth affective labels as determined within the context of interest. These labels are commonly obtained by ratings of affective experience as given by the individual whose experience is modelled, or as

given by other individuals (e.g. when using affective stimuli from standardized stimuli collections such as the affective picture collection IAPS [19]). However, as argued in the first paragraph, these ratings have certain properties (such as their conscious and possibly biased nature) that we want to avoid using implicit measures. We should not base the interpretation of implicit measures on exactly the measures whose properties we would like to avoid. In the current study, we therefore follow an approach that, in principle, does not require precise affective ratings. In this approach we train a model using the vinegar solution as an extreme high arousal, low valence stimulus, and then apply the model to the regular drinks. The vinegar solution may turn out to be an outlier that cannot be used to generalize to regular drinks. As described earlier, for some measures, this may be the case. However, if it does generalize sufficiently (as evidenced by an association between the model’s output score and self-reported ratings for regular drinks), this approach would be a promising way to gauge individuals’ affective experience when tasting food without needing to ask. As a comparison, we contrast this approach to a more traditional approach that does use the valence and arousal ratings as the ‘true’ affective experience of drinks.

2 METHODS

2.1 Participants

A total of 70 healthy participants (19 men, 51 women) took part in this study. They were between 19 and 63 years old, with an average of 48.5 years. All participants signed an informed consent in accordance with the Helsinki Declaration of 1975 as revised in 2014 [20], before participating in this study. The study was approved by the TNO Institutional Review Board (TCPE). Three participants were excluded due to technical problems related to the registration of event markers and physiological data recording. This left us with 67 participants for further analysis.

2.2 Materials

EEG, ECG and EDA were recorded using a Biosemi Active Two MkII system, with a sampling frequency of 512 Hz. For EEG, 32 active silver-chloride electrodes were placed according to the 10-20 system. ECG electrodes were placed on the right clavicle and on the lowest floating left rib. EDA was measured by placing gelled electrodes on the fingertips of the index finger and the middle finger of the non-dominant hand. Pupil diameter was recorded at 60 frames per second using SmartEyePro V6.1.6. This system consists of two cameras, placed at the left and right side of a screen that presented the name of the drink and the rating scales. The screen had a size of 37.0 by 30.0 cm and the viewing distance was approximately 80.0 cm. Room lightening was normal office lightening, and the same for all participants.

SAM pictures [21] with 9-point scales were used for valence and arousal self-report ratings. The 9-point scale was positioned in the appropriate location at the bottom of each SAM scale, where the most leftward (most unpleasant and calm) and the most rightward (most pleasant and aroused) parts of the scale were translated into values of 1 and 9,

respectively. Participants entered their rating by clicking on one of nine small circles.

For the behavioral measure of sip size, the exact weight of each drink including the cup was measured before the participant took a sip. After finishing the experiment, the cups with the rest of each drink were weighted again to determine the sip size in grams.

The drinks used in this study were apple juice, orange juice, yogurt drink, milk, buttermilk, rooibos tea, black tea, cola and a vinegar solution (50% vinegar, 50% water). The regular drinks were chosen to represent a variation in basic flavors and temperature. They clearly differed from one another in taste, but they were expected to be close to one another in affective experience, at least relative to diluted vinegar. Teas and the vinegar solution were prepared in the same way each morning. Teas were kept at about 60 °C, and the vinegar solution was kept at room temperature. The other drinks were kept in a refrigerator before being served to the participants. Each sample was served in a white plain cup, in portions of 50 g. Participants tasted the drinks in randomized order, except that the vinegar solution was always followed by a cup of water, to decrease possible lingering of emotional and physiological effects. Responses to water are therefore not included in the analyses.

2.3 Design and Procedure

After participants arrived at the laboratory, the experimental procedure was explained, and they were asked to sign the informed consent. The electrodes for EDA, ECG, and EEG were attached, and participants were asked to sit comfortably in front of a screen. Participants filled out a general questionnaire on demographic details and current emotional state. Before the experiment started, the SmartEyePro system was calibrated. Then, the experimenter showed and explained how to take a sip, immediately putting the cup down after the sip, and participants performed a practice trial with water. After this there was time for additional practice or instructions when needed, and participants had the chance to ask questions. The timeline of an experimental trial is indicated in Fig. 1. First, the name of drink was presented on the screen. This was the sign for the experimenter to place the appropriate drink in front of the participant. After 5 seconds, the name of the drink disappeared, which was the sign for the participant to take one sip. Participants were instructed and practiced to always make the same drinking movement. After taking the sip, participants sat still and looked at a blank white screen. Forty seconds after the name of the drink had appeared on the screen, the self-report valence and arousal rating scales appeared. After rating valence and arousal, the name of next drink appeared on the screen. This procedure was repeated until all ten drinks (eight regular drinks, vinegar solution and water) had been served once. Note that in our procedure, and as common in daily life, participants always knew what drink they were about to taste. The food experience and physiological processes that we study thus reflect a mixture of expectation and sensory processes, starting at the moment that the name of the drink appears on the screen.

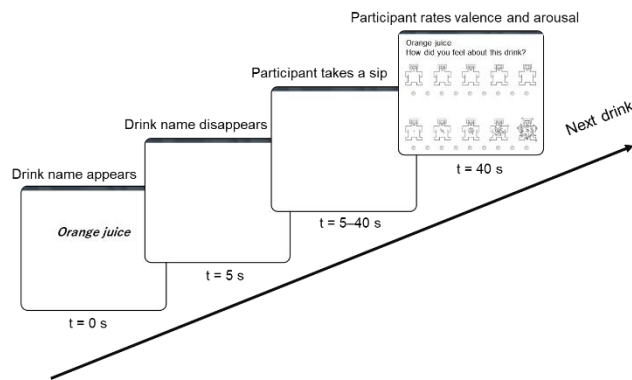


Fig. 1. Outline of an experimental trial. After participants rated the drink, the next trial started in which another drink was presented. In this manner, each participant responded to each of the ten different drinks once.

2.4 Data Processing and Analysis

Before further analysis, EDA, ECG and EEG were pre-processed.

The EDA signal was bandpass filtered between 0.03 and 100 Hz with custom made Matlab algorithms.

Inter-beat interval (IBI), defined as the temporal distance between R-spikes, was extracted from the ECG signal using custom made Matlab algorithms.

Raw EEG data were pre-processed and analyzed using MATLAB and the FieldTrip open source MATLAB toolbox [22]. The EEG pre-processing entailed standard procedures of referencing the signals to the average EEG signal and filtering them using a 0.5 Hz high pass and a 43 Hz low pass filter to remove slow drifts and high-frequency noise. Logistic infomax independent component analysis (ICA, [23]) was performed to classify artifactual independent components, i.e., components not reflecting sources of neural activity, but rather ocular or muscle-related artifacts. These components were removed from the data. This was done using EEGLAB v14.1.2 for MATLAB [24].

Valence and arousal ratings were normalized by subtracting the average rating of a participant of each of her/his ratings so that for each participant, the average valence and arousal rating was 0.

Table 1 gives an overview of all extracted features. Summing the number of features gives a feature vector length of 152 for each drink and each participant. The features were extracted as described in the following.

Sip size was log transformed to obtain more uniform distributions and normalized like the valence and arousal ratings.

The EDA signal was cut in 40s epochs, starting at the onset of the name of the drink and ending 40 seconds after, for each drink and each participant. Using a Matlab toolbox for analyzing EDA (Ledalab, [25]), several features reflecting phasic and tonic activity were extracted from the raw signal, where possible using both Continuous Decomposition Analysis (CDA) and Through To Peak (TTP) analysis. Values were normalized as described above for valence and arousal.

Several time series features were determined. For this, pupil size, EDA, IBIs and EEG were cut in eight 5s epochs,

starting at the onset of the name of the drink and ending 40 seconds later, at the time that the participants scored their experience. For each of these 5s epochs, features were extracted (e.g., the mean or median over these 5 seconds) to form a time series response. Note that 5s intervals are too short to quantify parasympathetic activity from heart rate variability [26]. Heart rate variability as computed here should thus be seen as an indicator of the extent to which successive IBIs change.

To arrive at EEG alpha asymmetry, the following steps were taken. For each of the intervals, the spectral power was calculated over bands ranging from 8 to 13 Hz (alpha) in steps of 0.2 Hz following a fast Fourier transform (FFT) approach using a single Hanning taper. Subsequently, values were integrated. FAA at F3 (F7) and F4 (F8) was determined for each 5s segment by taking the relative difference between alpha as recorded at the right and the left side of the cortex: $((R - L) / (R + L)) * 100$ [27]. Positive values indicate lower alpha power in the left than in the right hemisphere (i.e., relatively greater left hemisphere cortical activity). Besides alpha asymmetry, alpha at Pz and theta (4-8 Hz) at Fz were included as variables, since these might be informative of agitation or arousal [28], [29].

All time series variables, except for heart rate variability, were normalized per individual, drink and variable, by subtracting the series means from the series. In addition to the 8 values reflecting the subsequent 5s epochs, a 9th value reflecting the mean was added to form time series of nine values.

For each variable (feature), before and after normalization, values that were more than 5 standard deviations away from the mean were discarded as outliers. These missing data (<2%) were imputed using k-nearest neighbors imputation.

TABLE 1
FEATURES

Sensor category	Main variable	Variable version	Nr of features
Behavior	Sip size	Sip size	1
Pupil diameter	Pupil diameter time series	Median	9
		Standard deviation	9
EDA	Raw skin conductance	Mean	9
		Median	9
	level time series	Minimum and maximum of mean	2
		Minimum and maximum of median	2
	Phasic skin conductance (Ledalab)	Number of responses (CDA and TTP)	2
		Summed response amplitude (CDA and TTP)	2
		Response latency (CDA and TTP)	2
		Global mean	1
	Global maximum deflection	Global maximum deflection	1
		SCR: average response	1

	Phasic maximum	1	
	Time integral of response	1	
Tonic skin conductance (Ledalab)	Tonic	1	
	Inter beat interval time series	Mean Median	9 9
ECG	Heart rate variability time series	RMSSD	9
	Alpha asymmetry time series	F3F4	9
		F7F8	9
EEG	Alpha time series	F3	9
		F4	9
		F7	9
		F8	9
		Pz	9
Theta time series	Fz	9	

2.5 Modelling

For our approach to use the vinegar solution as ground truth high arousal, low valence; and the regular drinks as ground truth low arousal, high valence stimuli, we trained an elastic net binomial classification model to distinguish between vinegar and all other drinks. All implicit variables were used for this. The model used 10-fold cross validation to determine the optimal lambda and to obtain error rates. More specifically, the procedure ran for 11 (10-folds+1) times; the first time to get the lambda sequence, and then the remainder to compute the fit with each of the folds omitted. The average error and standard deviation over the folds were computed. Minimum average classification error was reached for a certain lambda which led to the final classification accuracy. Upsampling (random sampling with replacement) was used to deal with class imbalance (i.e., the fact that vinegar was a less frequent stimulus compared to regular drinks). The final error rate is reported for balanced classes.

In addition, combinations of variables were tested with the idea of choosing a limited, but still sensitive sensor set. For this, we tested all pairs of the 'sensor categories' behavior, pupil diameter, EDA, ECG and EEG.

For the comparison approach (using explicit valence and arousal as ground truth labels), an elastic net regression model was used using all implicit variables. Again, 10-fold cross validation was used to find the optimal lambda and to obtain error rates. For this approach, only data from regular drinks was used.

3 RESULTS

3.1 Approach 1: Using the Vinegar Solution as Ground Truth

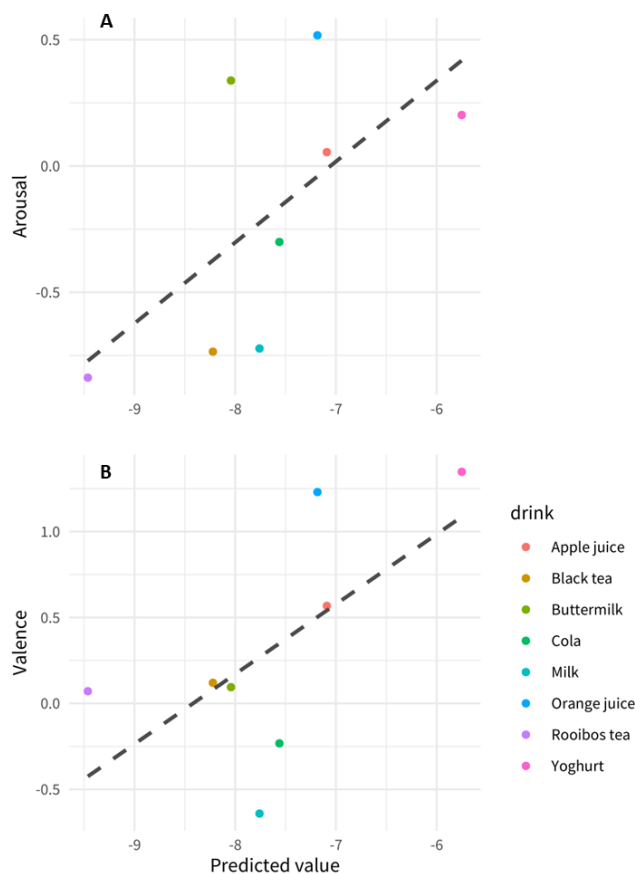


Fig. 2. Scores assigned by the model trained on all implicit features versus rated arousal (A) and rated valence (B), averaged across drinks.

Training a linear elastic net binomial classification model on all implicit variables using balanced classes ‘vinegar’ and ‘regular drinks’ resulted in a classification accuracy of 91%. The trained classification model was then used to assign continuous scores to the regular drinks. A high score means that the data is judged to be closer to vinegar (closer to high arousal, low valence); a lower score means that the data is judged closer to regular drinks (closer to low arousal, high valence). Indeed, we found a positive correlation between the mean drink scores as assigned by the model, and mean rated arousal (Pearson’s Rho $R=0.64$, $p=0.04$, Fig. 2A). We also found a trend of a positive, rather than negative, correlation between the mean predicted scores and mean rated valence ($R=0.63$, $p=0.05$, Fig. 2B). As a reference, the model’s assigned score for vinegar was 3.6. Rated arousal for vinegar was 1.0, and rated valence -2.4.

The best combination of any two implicit sensor categories was ECG and sip size (which was in accordance with our finding that the top 23 largest weights in the general model were ECG derived features; sip size ranked 43th of the 152 features). Training the model using features from only ECG and sip size resulted in a classification accuracy of 76% distinguishing vinegar from regular drinks. Similar

as what we found when using the model above, mean drink scores as predicted by this model correlated positively with mean rated arousal ($R=0.87$, $p<0.01$, Fig. 3A). No correlation with valence was found ($R=0.46$, $p=0.13$, Fig. 3B). The score of vinegar in this model was 0.8.

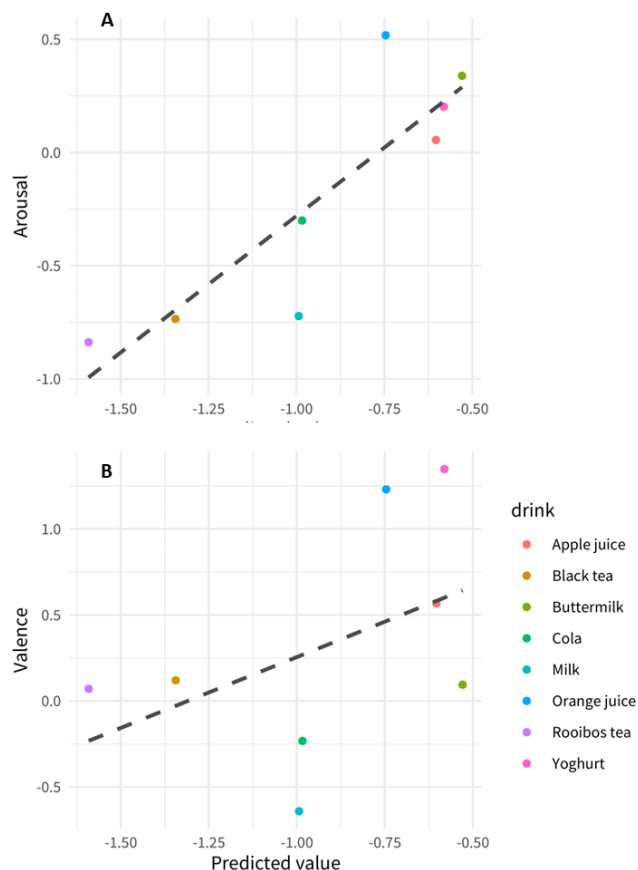


Fig. 3. Scores assigned by the model trained on features derived from ECG and sip size versus rated arousal (A) and rated valence (B), averaged across drinks.

3.2 Approach 2: Using Explicit Valence and Arousal as Ground Truth

Training an elastic net regression model using all implicit variables to predict arousal ratings and valence ratings for all regular drinks was not successful. Model performance as judged by RMSE values obtained during the cross-validation procedure was very low. The model reached best performance (i.e. lowest RMSE values) with variable coefficients of 0; this means that any linear combination of predictors will result in worse performance compared to simply using the mean rated arousal or valence of all drinks as the predicted value.

4 DISCUSSION

In the current study, we explored the potential of combining implicit measurements, covering the behavioral, physiological and neurophysiological domain, to estimate affective experience when tasting drinks.

Training a model using sip size, pupil diameter, ECG, EDA, and EEG using the vinegar and regular drinks data performed well in attributing unseen data to either of these

two categories (91% correct). Thus, it is possible to estimate from implicit data from an individual tasting a particular drink, whether this individual tasted a regular drink or an unpleasant, high arousal drink. When limiting the number of sensor categories to two, ECG and sip size are the best choice, reaching a classification accuracy of 76%. Relatively high sensitivity of heart rate and sip size to distinguish vinegar from regular drinks is consistent with our earlier analysis of these data [17].

For most potential applications, models should be able to estimate affective experience for stimuli that differ more subtly in affect than the vinegar solution and regular drinks. When we only used the implicit data obtained from individuals tasting regular drinks, and used rated arousal and valence as ground truth affective experience, it appeared to be impossible to estimate rated arousal and valence from unseen data. However, when we used the model described above, trained by only using implicit data obtained from individuals tasting regular drinks as well as a known ground truth high arousal low valence drink (vinegar), and applied this model to regular drinks, model scores correlated with rated arousal in the direction as expected. For valence, there was an unexpected positive correlation trend, where a high score ("similar to vinegar") tended to correspond to high rather than low valence. As discussed and substantiated in [17], we think that this is due to the fact that in regular drinks, rated valence and arousal correlate positively. Most of our implicit measures reflect arousal, such that the model distinguishing between vinegar and regular drinks capitalizes on the difference in arousal rather than valence. Together this leads to the observed positive trend in the correlation between predicted score and valence in regular drinks.

Our results indicate that implicit responses to a quite extreme, affective ground truth stimulus can be generalized to responses to stimuli that differ more subtly in affective experience in the case of tasting drinks. Starting with affective ratings as ground truth (as is usually done) did not work out in this case, possibly because the ratings were too variable and noisy. Note that, in order to validate our approach, we did make use of affective ratings (showing the correlation between rating and model score), but in principle, this is not required.

As found before, our results indicated that implicit measures generally better match arousal than valence. Most (food related) research focusses on measures reflecting pleasantness. At first sight, this may seem the more interesting measure, but if one is interested in stimuli of which it is a priori known that they are all relatively pleasant (as is the case with most food stimuli), arousal may be of special interest. This is especially so since arousal affects the memorability of an event [30], [31]. In [32] and [33] arousal is discussed as a crucial determinant in defining (sustained) attractiveness of products. In addition, explicit measures of arousal have been found to be vulnerable to the exact formulation of the question [34], which renders implicit measures all the more useful.

One may have expected that for arousal, EDA would have turned up as the most important predictor. Sweat glands are exclusively innervated by the sympathetic

branch of the Autonomic Nervous System [35] and EDA has been associated consistently with arousal across a range of contexts, whereas the link between heart rate and arousal is less clear. While associations between heart rate and arousal are usually positive, negative associations have been found as well (e.g. within the context of reading a novel [36]). However, working against EDA is the large variability between individuals (e.g. [37]), which may have resulted in our finding that within the context of tasting drinks, ECG related variables are more suitable to estimate arousal than EDA.

As explained in the introduction, we chose for a design resembling real-life, natural tasting of drinks. This resulted in certain limitations, such as added noise and being unable to disentangle expectation and sensory effects on food experience and physiological processes. We also do not know to what extent the pure effect of chemical properties of the drinks (or more precisely, their interaction with the senses of an individual) are responsible for the results. For instance, perceived sourness is one of the properties that differs between vinegar and the regular drinks, and the model may have captured that instead of arousal. If arousal and perceived sourness are associated in regular drinks, this could explain the association between rated arousal and the model's score. To get an impression of the importance of this possibility, we trained a linear model to use implicit measures simply for distinguishing between the different drinks. If implicit measures respond relatively strongly to more basic, chemical properties rather than affect, this should be possible. However, this model's classification accuracy (5%) did not exceed chance level (which is 12.5%), suggesting that differences in chemical composition of our drinks do not result in clear differences in implicit responses.

For future research, including multiple types of ground truth stimuli would be valuable. Including a stimulus that is a priori (i.e., for everyone) known to be high arousal - high valence relative to the regular stimuli would be good, but in the case of food stimuli hard to do because regular food stimuli are all high in valence. Still, including multiple types of high arousal - low valence stimuli (e.g. a quinine solution next to the vinegar solution in the present study) is possible and may result in better generalizability and robustness. While this study showed that we already can generalize a model trained on distinguishing a vinegar solution from regular drinks to estimating affective experience in regular drinks, this was shown only for rated arousal and not for rated valence. Also, the finding that the model that distinguished the two classes best (91% accuracy for the model using all features versus 76% for the model using ECG and sip size) did not result in the highest correlation between model scores and rated arousal for regular drinks (cf. Fig 2A and 3A), indicates that generalizability was not perfect. This may improve when multiple instances of ground truth stimuli are included.

7 CONCLUSION

We demonstrated that food experience (arousal) elicited by a range of regular drinks could be estimated using

implicit measures without first having to rely on the measures that we aim to avoid, namely, ratings of affective experience. Estimates of affective experience purely based on implicit, ‘unconscious’ responses may be helpful to understand consumers’ food-evoked emotions better compared to only using self-reports. Future research should focus on improving the model e.g. by using different ground truth stimuli, and on relating scores resulting from models based on implicit measures, as well as explicit measures, to variables of interest such as future food choice behavior.

ACKNOWLEDGMENT

The authors wish to thank Esther Bosch, Wouter Korteling, and Benjamin de Graaff for their assistance with conducting the experiment. This work was supported by Kikkoman Europe R&D Laboratory B.V. Corresponding author is Anne-Marie Brouwer.

REFERENCES

- [1] Prescott, J. (2017). Some considerations in the measurement of emotions in sensory and consumer research. *Food Quality & Preference*, 62, 360-368.
- [2] Kaneko, D., Toet, A., Brouwer, A.-M., Kallen, V., & van Erp, J. B. F. (2018). Methods for evaluating emotions evoked by food experiences: A literature review. *Frontiers in Psychology*, 9, 911.
- [3] Lagast, S., Gellynck, X., Schouteten, J. J., De Herdt, V., & De Steur, H. (2017). Consumers’ emotions elicited by food: A systematic review of explicit and implicit methods. *Trends in Food Science & Technology*, 69, 172-189.
- [4] Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., et al. (2015). Predicting Advertising Success Beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling. *Journal of Marketing Research*, 52(4), 436-452.
- [5] Gutjar, S., de Graaf, C., Kooijman, V., de Wijk, R.A., Nys, A., ter Horst, G. J., et al. (2015). The role of emotions in food choice and liking. *Food Research International*, 76, 216-223.
- [6] Toet, Kaneko, Ushiyama, Hoving, de Kruijf, Brouwer, Kallen, Van Erp (2018) EmojiGrid: a 2D pictorial scale for the assessment of food elicited emotions. *Frontiers in Psychology*, 9, 2396
- [7] Köster, E. P., & Mojet, J. (2015). From mood to food and from food to mood: A psychological perspective on the measurement of food-related emotions in consumer research. *Food Research International*, 76, 180-191.
- [8] Verastegui-Tena, L., Schulte-Holierhoek, A., van Trijp, H., & Piqueras-Fiszman, B. (2017). Beyond expectations: The responses of the autonomic nervous system to visual food cues. *Physiology & Behavior* 179:478-486.
- [9] Danner, L., Sidorkina, L., Joechl, M., & Duerrschmid, K. (2014). Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology. *Food Quality and Preference*, 32, 167-172.
- [10] de Wijk, R. A., Kooijman, V., Verhoeven, R. H. G., Holthuysen, N. T. E., & de Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. *Food Quality and Preference*, 26(2), 196-203.
- [11] Rousmans, S., Robin, O., Dittmar, A., & Vernet-Maury, E. (2000). Autonomic nervous system responses associated with primary tastes. *Chemical Senses*, 25(6), 709-718.
- [12] de Wijk, R. A., He, W., Mensink, M. G. J., Verhoeven, R. H. G., & de Graaf, C. (2014). ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks. *PLoS One*, 9(4, paper e93823), 1-9.
- [13] Mendes, W. B. (2009). Assessing autonomic nervous system activity. In J. S. Beer, & E. Harmon-Jones (Eds.), *Methods in social neuroscience* (pp. 118-147). New York, NY: Guilford Press.
- [14] Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological Psychology*, 84(3), 451-462.
- [15] Gable, P., & Harmon-Jones, E. (2008). Relative left frontal activation to appetitive stimuli: Considering the role of individual differences. *Psychophysiology*, 45(2), 275-278.
- [16] Brouwer, A.-M., Hogervorst, M., Grootjen, M., van Erp, J., & Zandstra, E. (2017). Neurophysiological responses during cooking food associated with different emotions. *Food Quality and Preference*.
- [17] Kaneko, D., van Erp, J.B.F., Hogervorst, M.A., Toet, A., Kallen, V. & Brouwer, A.-M. (2019) Physiological responses to tasting drinks associated with different tasting experiences. *Sensors*, 19(20), 4397.
- [18] Brouwer, A., Zander, T., van Erp, J.B.F., Korteling, J., & Bronkhorst, A. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience*, 9.
- [19] Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture System (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL
- [20] World Medical Association (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA*, 310, 2191-2194.
- [21] Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
- [22] Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011, 1.
- [23] Bell, A.J., Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7, 1129-1159.
- [24] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134, 9-21.
- [25] Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190, 80-91.
- [26] Task Force of the European society of cardiology and the North American society of pacing and electrophysiology (1996). Heart Rate Variability – Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17, 354-381.
- [27] Papousek, I., Weiss, E. M., Schuster, G., Fink, A., Reiser, E.M., & Lackner, H. K. (2014). Prefrontal EEG alpha asymmetry changes while observing disaster happening to other people: Cardiac

correlates and prediction of emotional impact. *Biological Psychology*, 103, 184-194.

- [28] Pfurtscheller, G., Stancak, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band: an electrophysiological correlate of cortical idling. *International Journal of Psychophysiology*, 24, 39-46.
- [29] Bekkedal, M.Y.V., Rossi, J., & Panksepp, J. (2011). Human brain EEG indices of emotions: Delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neuroscience & Biobehavioral Reviews*, 35(9), 1959-1970.
- [30] Anderson, A. K., Wais, P. E., & Gabrieli, J. D. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences*, 103(5), 1599-1604.
- [31] McGaugh, J.L. (2006). Make mild moments memorable: add a little arousal. *Trends in Cognitive Sciences*, 10 (8), pp. 345-347.
- [32] Köster, E. P., & Mojet, J. (2006). Theories of food choice development. In L. J. Frewer, & J. C. M. Trijp (Eds.), *Understanding consumers of food products* (93 - 124). Cambridge, United Kingdom: Woodhead Publishing.
- [33] Köster, E. P., & Mojet, J. (2007). Boredom and the reasons why some new products fail, in H. MacFie (Ed.), *Consumer-Led Food Product Development* (pp. 262-280). Cambridge, United Kingdom: Woodhead Publishing Limited
- [34] Toet, A., Kaneko, D., Ushiyama, S., Hoving, S., de Kruijff, I., Brouwer, A.-M., Kallen, V., & van Erp, J.B.F. (2018). EmojiGrid: a 2D pictorial scale for the assessment of food elicited emotions. *Frontiers in Psychology*, 9, 2396.
- [35] Dawson, M.E., Schell, A.M., & Filion, D.L. (2016). The Electrodermal System. In *Handbook of Psychophysiology*, 4th ed.; Cacioppo, J.T., Tassinari, L.G., Berntson, G.G., Eds.; Cambridge Handbooks in Psychology, Cambridge University Press: Cambridge, UK.
- [36] Brouwer, A.-M., Hogervorst, M. A., Reuderink, B., van der Werf, Y., van Erp, J. B. F. (2015). Physiological signals distinguish between reading emotional and non-emotional sections in a novel. *Brain-Computer Interfaces* 2(2-3), 76-89.
- [37] Thammasan, N., Stuldreher, I. V., Schreuders, E., Giletta, M., & Brouwer, A.-M. (2020). A usability study of physiological measurement in school using wearable sensors. *Sensors*, 20(18), 5380.



Anne-Marie Brouwer received her PhD degree from the Erasmus University Rotterdam in The Netherlands, in 2002. She was a postdoc studying visuomotor coordination at the Max Planck Institute, Tübingen, Germany, and Rochester University, New York, NY. She is now Senior Scientist at The Netherlands Organization for Applied Scientific Research TNO. Since 2007 Anne-Marie has been working on (possible applications of) physiological correlates of mental state in a range of contexts. She is currently Specialty Chief Editor of *Frontiers in Consumer Neuroergonomics* and active in a number of review and conference boards.



Tim J. van den Broek received his Master's degree in Pharmaceutical Sciences from Utrecht University. He is currently working as a data scientist at The Netherlands Organization for Applied Scientific Research in the department of Microbiology & Systems Biology. He is currently working primarily on multi-modal data integration in the domain of personalized health.



Maarten A. Hogervorst received his PhD in physics from the University of Utrecht in 1996. After a post-doc position at Oxford University he started working at The Netherlands Organization for Applied Scientific Research TNO. His current research interests include human visual information processing, electro-optical system performance, search and target acquisition modeling & camouflage, image quality assessment, image enhancement and mental state estimation from (electro)physiology. Maarten is senior member of the SPIE.



Daisuke Kameko (BSc in University of Oregon in US, MSc in University of Tsukuba in Japan) is affiliated with Kikkoman Europe R&D Laboratory B.V. since February 2016 as a R&D manager. He is now also an external PhD candidate in University of Twente in the Netherlands. His current research interests include evaluating food-evoked emotions and effects of food experience on food preference with implicit measures, such as physiological and behavioral measures.



Alexander Toet received the PhD degree in physics from the University of Utrecht, The Netherlands, in 1987. He is currently affiliated with The Netherlands Organization for Applied Scientific Research TNO and with the Department of Experimental Psychology of Utrecht University. His current research interests include multisensory and affective perception, visual saliency, clutter and multimodal image fusion. He is a fellow of the SPIE and a senior member of the IEEE.



Victor Kallen received his PhD in Medicine from Erasmus University / ErasmusMC in Rotterdam. He used to be senior scientist at The Netherlands Organization for Applied Scientific Research TNO, Healthy Living, and is nowadays senior scientist at TNO Defense, Safety and Security. His primary scientific interests are in the arena of personalized predictive models to quantify individual health and/or performance prognoses. He is member of the International Society of Psychoneuroendocrinology (ISPNE).



Jan van Erp is principal scientist with The Netherlands Organization for Applied Scientific Research (TNO) and full professor of tangible user interaction with the University of Twente. Jan obtained a master degree in Experimental Psychology from Leiden University and a PhD in Computer Science from Utrecht University, both in The Netherlands. His research focuses on multisensory perception and cognition, applied neuroscience, robotics, and human-machine collaboration in complex environments. Jan published more than a hundred peer reviewed papers (h-index 38) and holds several auxiliary board and advisory functions and serves on the editorial board of four scientific journals. Jan is the current vice-president of the Eurohaptics Society, board member of the foundation ICT for Brain, Body and Behaviour, general chair of Eurohaptics 2020, chair of the TNO society of principal scientists, and chair of the TNO Institutional Review Board.