Food Quality and Preference 68 (2018) 349–359

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Automated facial expression analysis for emotional responsivity using an aqueous bitter model

C.A. Crist[a], S.E. Duncan[b,*], E.A. Arnade[b,1], K.A. Leitch[b,2], S.F. O'Keefe[b], D.L. Gallagher[c]

[a] Department of Food Science, Nutrition and Health Promotion, 945 Stone Boulevard, Mississippi State University, Mississippi State, MS 39760, USA
[b] Department of Food Science and Technology, 1230 Washington St. SW, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA 24061, USA
[c] Department of Civil and Environmental Engineering, 1145 Perry Street, Virginia Tech, Blacksburg, VA 24061, USA

ARTICLE INFO

Keywords:
Emotion
Facial expression analysis
Bitter
Caffeine
Sensory evaluation
Time series

ABSTRACT

The food industry is seeking ways to understand consumer emotions, using implicit measurements, to differentiate acceptability of products in the marketplace. Automated facial expression analysis (AFEA) is a prospective analysis for product acceptability. This study used aqueous bitter solutions to determine and validate AFEA as an analysis supplement to product liking. Participants (n = 46) evaluated a control (distilled water) and three bitter (caffeine) solutions: low (0.05% w/v); medium (0.08% w/v); and high (0.15% w/v). Individual participant sessions were video-recorded and analyzed (5 s; $\alpha = 0.20$) for each sample in the default and continuous analysis setting. Participants rated liking and bitter intensity on a 9-point scale. An inverse relationship existed between liking and bitter intensity ($r_s = -0.90$; $p < 0.0001$). In continuous setting for AFEA analysis of mean emotion intensity, analyzed by ANOVA, only the medium bitter treatment elicited a higher disgust response control ($p < 0.20$) and no differences were found between treatments in disgust ($p > 0.20$) evaluations using program default settings. For time series analysis with both the continuous and default settings, disgust was a predominant emotion in the medium and high bitter solutions as well as happy in the high ($p < 0.025$). Using time series analysis, continuous and default results had similar patterns over 5 s, but continuous data was more intermittent. Time series analysis is a promising tool for interpreting emotional results of a population and is more sensitive to emotional changes than mean comparisons. Future studies should continue to improve the characterization and sensitivity of emotions to food acceptability using AFEA.

1. Introduction

Automated facial expression analysis (AFEA) application, as a tool for assessing consumer emotional response to food stimuli, is in its infancy, with limited published literature (Supplementary Material (SM) Table 1). Motivation for characterizing consumers' emotions toward foods and beverages is largely associated with predicting acceptability, consumer choice, and consumption behavior. Beverages are commonly used products in the available AFEA literature, including commercial breakfast drinks (De Wijk, He, Mensink, Verhoeven, & de Graaf, 2014), juices (Danner, Sidorkina, Joechl, & Duerrschmid, 2014; Danner, Haindl, Joechl, & Duerrschmid, 2014), basic taste solutions (Wendin, Allesen-Holm, and Bredie, 2011; Arnade, 2013; Zhi, Cao and Cao, 2017), and milk (Arnade, 2013; Walsh, Duncan, Potts and Gallagher, 2015). Most authors suggest that facial expressions provide additional information for differentiating among samples and/or characterizing acceptability of a stimulus. Zhi et al. (2017), using Asian participants and basic taste stimuli, determined that bitter stimulus was associated with more negative emotions.

Interpretation of AFEA-classified emotions in relation to product acceptability is difficult and often based on averaging of universal emotion intensity ratings over time and across participants. In a study using high and low concentrations of compounds eliciting basic tastes, Arnade (2013) found, in both high and low concentration sessions, that the mean for sad emotion, averaged over the tested time interval, was higher than that of the angry, scared, disgusted, or happy emotions. The differences in emotion characterization among basic tastes were not as great as expected, thus questioning the accuracy of current methods for emotional capture or their statistical analysis (Arnade, 2013). Walsh et al. (2015) documented that variability in AFEA responses to fluid

Abbreviations: AFEA, automated facial expression analysis

* Corresponding author at: 104 Hutcheson Hall, 250 Drillfield Dr., Virginia Tech, Blacksburg, VA 24061, USA.
E-mail addresses: cac400@msstate.edu (C.A. Crist), duncans@vt.edu (S.E. Duncan), dang@vt.edu (D.L. Gallagher).
[1] Current location: Pilgrim's, Greeley, CO, USA.
[2] Current location: McCormick & Co., Inc. Hunt Valley, MD, USA.

https://doi.org/10.1016/j.foodqual.2018.04.004
Received 11 September 2017; Received in revised form 7 April 2018; Accepted 9 April 2018
Available online 12 April 2018
0950-3293/

milk off-flavor was high and difficult to interpret based on summary data expressed as both means and through overall proportion of universal emotions as illustrated by pie chart analysis. Arnade (2013) found high variability among individual emotional response to chocolate milk and white milk. However, even with this variability, panelists elicited a happy response from samples longer than sad and disgusted (Arnade, 2013). In the first reported use of time series analysis for AFEA data associated with foods, Leitch, Duncan, O'Keefe, Rudd, & Gallagher (2015) found temporal trends of emotions relating to natural and artificial sweeteners in a cold tea beverage. The limited number of food and beverage-related studies and diversity in approaches used for assessing AFEA data creates challenges in interpreting sensitivity of the automated software and interpreting implicit emotions associated with the food and beverage stimuli. Methodological guidance to enhance data analysis and interpretation for AFEA are needed as this research field expands.

There are several commercially available AFEA software systems, summarized in SM Table 1, that can characterize emotions elicited from a stimulus; however, most peer reviewed published research has used FaceReader™ (Noldus Information Technology, Wageningen, The Netherlands) for food and beverage consumption and emotional analysis (SM Table 1). Use of AFEA software allows for dynamic evaluation over time but analyses are still challenged by limited understanding of appropriate data interpretation relevant to food and beverage stimuli. At present, no standard methodology for food and beverage consumption analyses exists, as illustrated by the broad range of research methodology reported in the studies published using FaceReader™ (SM Table 1). Many psychology studies of emotions utilize intense stimuli (Davidson, Ekman, Saron, Senulis, & Friesen, 1990), which may be more readily captured by manual facial action coding units system (FACS) and AFEA; such studies are the framework for most algorithms and recommendations related to AFEA analyses. In most cases, food does not elicit intense emotional responses (Walsh, Duncan, Bell, O'Keefe, & Gallagher, 2017b), which may require further evaluation before a calibration recommendation approach is made.

Appropriate methodological approaches for AFEA application to foods and beverages is contingent on selection of sensitive and logically accurate software settings to capture subtle changes in facial expression (microexpressions) and appropriate analyses for finding overall and temporal differences. AFEA software may include multiple analysis settings. FaceReader™ 6 (Noldus Information Technology, 2014b) offers three analysis settings: default (no calibration), continuous calibration, and individual calibration. For the individual calibration in FaceReader™ 6, a participant's neutral expression is used. Continuous calibration consists of software actively eliminating participant expression bias (i.e. some people look sad by nature; data capture setting is not optimal) while running analysis without individual calibration images or video (Noldus Information Technology, 2014b). It was suggested by a Noldus representative to use continuous calibration with food product evaluation (A. Macbeth, personal communication, February 15, 2015). Calibration settings of the software should be considered to improve sensitivity and accuracy of interpretation but guidance pertaining to food and beverage applications is not yet reported.

There is no consensus on optimal methodology to statistically analyze and interpret output. Crist, Duncan, and Gallagher (2016) described a protocol for data collection and analysis for AFEA and temporal analysis of beverages and soft foods to assist in procedural and statistical design. Development of AFEA methods for implicit emotional response to foods and beverages might improve understanding of consumer affective response. Such a tool may provide a more unique and deeper relationship with brands and its consumers. This deeper connection has the potential to improve overall consumer experience and emotional investment. Our goal was to use a simple stimulus, common in foods and beverages, with a known facial expression response (caffeine; bitter) to develop a basis for enhancing sensitivity and accuracy of interpretation, and to propose recommendations for improved AFEA assessment of food and beverages.

In the assessment of bitter solutions using AFEA, this study evaluated:

A. Consumer liking:
   1. Consumer liking of aqueous bitterness solutions (caffeine) using hedonic ratings;
   2. Consumer liking as it relates to facial expressions associated with universal implicit emotions, as identified and measured by AFEA. Hypothesis: With increasing concentrations of a bitter compound, the hedonic response would decrease and disgust facial expression would increase.

B. AFEA calibrations and analysis settings for optimizing assessment:
   1. Analysis of AFEA videos using default and continuous calibration settings to determine a recommendation for application to beverage analysis;
   2. Evaluation of the sensitivity of the calibration settings appropriate for beverage analysis.Hypothesis: Continuous calibration setting would provide higher sensitivity to subtle changes (microexpressions)
      in facial expression in the context of this study.

C. Statistical analysis using time series for characterizing AFEA temporal differences:
   Hypothesis: Time series analysis of emotion states would provide detailed emotional analysis and results that differentiate products over time

2. Materials and methods

2.1. Sample preparation

Aqueous bitter treatment solutions were prepared as described by the Spectrum™ Descriptive Analysis Method (Meilgaard, Civille, & Carr, 2007) using caffeine (Sigma Aldrich, St. Louis, MO) in distilled water (The Kroger Co., Cincinnati, OH). We targeted four bitter intensity levels: control (distilled water); low (Spectrum™ 2; 0.05% (0.5 mg caffeine/mL distilled water) solution in water); medium (Spectrum™ 5, 0.08% (0.8 mg caffeine/mL distilled water); and high (Spectrum™ 10, 0.15% solution in water (1.5 mg caffeine/mL distilled water). The Spectrum™ Descriptive Analysis Method Intensity Scales Values (0–15) provide a standard reference for product evaluation using scaled intensities (Meilgaard et al., 2007). Solutions were poured into 2 oz. plastic sample cups (Monogram Company, Columbia, MD) and capped with color coded lids for ease of visual identification.

2.2. Consumer sensory analysis

2.2.1. Participant recruitment

The study was pre-approved by Virginia Tech IRB (IRB 13-037) prior to project initiation. Study recruitment was accomplished through email listservs to Virginia Tech faculty, staff, students and visitors. Potential participants were screened for this study after completing a bitter evaluation test. In the screening, participants tasted four bitter samples and those who rated the sample intensities in the appropriate order of increasing bitter concentration were identified as minimal risk for bitter blindness and included for recruitment in this study. Recruited participants completed a screening survey for personal attributes and demographics. Exclusion criteria included report of facial hair, required use of glasses for vision, allergies, bitter blindness, and age less than 18. Selected participants (n = 65; 18 male; 47 female; age range: 18–70) were Virginia Tech faculty, staff, students or visitors. Before sample evaluation, participants reviewed or consented to the study parameters, including video recording, before receiving additional instructions or samples.