International Journal of Industrial Ergonomics 66 (2018) 119–129

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

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

Comparing measurements for emotion evoked by oral care products

Yue Chen [a], Qin Gao [a, *], Qiaolu Lv [b], Nan Qie [a], Liang Ma [a]

[a] Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China
[b] Procter & Gamble Technology (Beijing) Co., Ltd., No.35 Yu'an Road, B Zone, Tianzhu Konggang Development Zone, Shunyi District, Beijing, 101312, P R China

ARTICLE INFO

Article history:
Received 9 March 2017
Received in revised form
7 December 2017
Accepted 20 February 2018
Available online 5 March 2018

Keywords:
Emotion
Flavor
Oral care products
Electroencephalography
Electromyography
Heart rate variability

ABSTRACT

The flavor of foods or oral care products can affect consumers' emotions and experience. We compared different methods for measuring emotion evoked by flavors, including self-report measures (Self-Assessment Manikin, or SAM and EsSense), electroencephalography (EEG), electromyography (EMG), and cardiovascular measures (HR and HRV). The results indicate that the difference of α/β power spectral density (PSD) ratios at AF4 and AF3 EEG channels can reflect emotion valence and produce the most consistent result for the 3 repetitions of the same stimulus. P8 β PSD and HR are reliable and valid for measuring emotion arousal. The two self-report measures, Self-Assessment Manikin (SAM) and selected items in EsSense Profile, can distinguish emotion evoked by five flavors. The divergent validity of self-reporting measures, however, is inadequate, which may be attributed to the halo effect, i.e., the strong perception of one emotional property influences people's perception of other emotional properties.

1. Introduction

With the rapid proliferation of new products, customers' emotional responses to a product has been found more and more important to the success of product success (Helander, 2001; Hsiao and Chen, 2006; Khalid, 2004). For oral care products, the flavor has a direct impact on emotional responses of customers. The sense of flavor is related to both the taste and the smell sensation. The chemicals in these products stimulate the taste buds and smell receptors, and the signals are sent to the insular cortex and the olfactory bulb in the brain. Close to the insular cortex and the olfactory bulb, an area involving emotions, called the amygdala, also receives signals. In this way, specific emotions are evoked by different flavors. Flavor-evoked emotions can further influence consumers' decisions to purchase oral care products (Damasio, 2006). The manipulation of product flavors is therefore an important way through which designers elicit desired emotional response from consumers.

To design product flavors that evoke specific positive emotions, being able to measure flavor-evoked emotions is an important step. Emotions, however, are complicated and difficult to measure accurately. So far, self-report measures are the most widely used tool. They are good for assessing mixed emotions and gathering rich interpretable opinions from consumers at low costs (Desmet, 2003; Paulhus and Vazire, 2007). To depict an accurate and comprehensive picture of customers' emotional response to a flavor, however, self-reporting measures have a couple of limitations. First, self-report measures cannot measure the emotion at exactly the moment when the emotion is evoked, whereas flavor-evoked emotion decays in seconds. Second, flavor stimuli can evoke unconscious or subtle emotion changes. Such subtle emotions may induce subliminal facial expressions, activate amygdala and other brain areas, and evoke skin conductance responses. But it is difficult to measure these by self-reporting because customers are not conscious of them (Berridge and Winkielman, 2003). Third, the results from self-reporting methods are affected by individual characteristics, such as cultural backgrounds (Desmet, 2003), one's the ability of reading and comprehension, and the ability to detect and be aware of one's emotions (Lane et al., 1997; Mauss and Robinson, 2009). These factors may confound self-reporting results. In addition, people may not be able to accurately describe their emotional feelings or they may deliberately modify their opinions if they do not want to express their true feelings, feel inhabited, or are unconsciously influenced by the circumstance, e.g., the experimenter and the design of questions (Mauss and Robinson, 2009; Paulhus and John, 1998; Czerwinski et al., 2001;

* Corresponding author.
E-mail address: gaoqin@tsinghua.edu.cn (Q. Gao).

https://doi.org/10.1016/j.ergon.2018.02.013
0169-8141/© 2018 Published by Elsevier B.V.

Nielsen and Levy, 1994).

To seek alternative measurement of emotions, a number of researchers have endeavored to develop emotion measures based on people's psychophysiological responses to the product (Guo et al., 2016; Hill and Bohil, 2016; Laparra-Hernández et al., 2009; Lee and Cho, 2009; Liu and Sourina, 2012; Peck et al., 2013; Qie et al., 2017; van den Broek and Westerink, 2009). The underlying rationale is straightforward: psychophysiological measures directly access people's primary response to an emotional stimulus without involving conscious processes (Kramer, 2006; Motte, 2009; Trimmel et al., 2009). By monitoring directly psychophysiological responses, we may infer about their emotional state. A major advantage of psychophysiological measures is that people cannot easily control their physiological signals voluntarily. In addition, they provide a continuous and real-time description of consumers' internal state, which is not possible with self-reporting methods, which are in nature retrospective.

Commonly used physiological measures for emotions include central nervous system measures (e.g., electroencephalography and neuroimaging), peripheral nervous system measures (e.g., skin conductance responses, heart rate, and heart rate variability), and facial measures (e.g., facial expressions and electromyography) (Mauss and Robinson, 2009). The feasibility of using these physiological methods to measure flavor-related emotions has been explored by a number of researchers (Brown et al., 2012; Hu et al., 1999; Park et al., 2011). These studies, however, focused on one to two specific physiological measures, used different flavor stimuli, and adopted different criteria for assessing the effectiveness of measurements. Thus, there is not a common ground for comparing these measures in terms of (1) the sensitivity to the difference in flavors used in foods and oral products, (2) the reliability to produce consistent results for the same stimulus, and (3) the validity to measure flavor-related emotions. Such knowledge is useful for flavor designers to choose suitable measures when evaluating products.

To address this void, this study compared self-report measures, EEG (electroencephalography), EMG (electromyography), heart rate (HR), and heart rate variability (HRV) in terms of their capability to measure emotion evoked by flavors. We collected 24 participants' emotional response to five flavors that are common in oral care products. The EEG, EMG, HR, and HRV were evaluated in terms of (1) the sensitivity to distinguish emotions evoked by flavors, (2) the reliability, and (3) the validity to reflect flavor-evoked emotions. Furthermore, by incorporating physiological measures, we attempted to develop an integrative model that can predict consumers' overall attitude towards and purchase intention of oral products.

2. Literature review

2.1. Theoretical model of emotions

A widely used model to describe emotions is the valence and arousal model proposed by Russell (1979). This model depicts emotions from the perspectives of valence (the direction of behavioral activation associated with emotion, either toward (positive) or away from (negative) a stimulus), and arousal (the extent or amount of physical response, from low to high). For example, the emotion of "happiness" is characterized by positive valence and high arousal. Using this model, however, "surprise" is also defined as a positive valence and high arousal emotion, though the two emotions are largely different. Mehrabian (1980) expanded this model by adding a dominance dimension (a feeling of being in control to a feeling of being controlled) and proposed the valence-arousal-dominance (VAD) mode. Using this model, "happiness" is defined as a positive, high arousal, high dominance emotion, whereas "surprise" is defined as a positive, high arousal, low dominance emotion.

2.2. Measures of emotion

Emotion measures fall into mainly two categories: self-report measures and physiological measures. By building upon VAD models, researchers have developed a number of emotion instruments, such as the Self-Assessment Manikin (SAM) (Hodes et al., 1985) and PAD (pleasure arousal dominance) emotion scales (Mehrabian, 1995). They assess emotions from the three state dimensions — pleasure (valence), arousal and dominance. Other emotion instruments, such as the Geneva Emotions Wheel (Scherer, 2005) and the Product Emotion Measurement instrument (PrEmo) (Desmet, 2003), measure emotional responses by directly specifying emotions such as "desired", "inspired", "satisfied", and "bored". In addition, researchers from different product domains developed specific instruments for typical emotions evoked by a specific type of stimuli, such as GEMS-25 (Zentner et al., 2008) for music-evoked emotions and ScentMove (Porcherot et al., 2010) for odor-evoked emotions. To measure emotion evoked by flavors, King and Meiselman developed the EsSense Profile (2010), which contains 39 terms (35 positive and 4 negative terms, such as "pleasant" and "disgusted"). The EsSense Profile has been widely applied in measuring emotion evoked by flavors of foods and beverages (de Wijk et al., 2012; Ferrarini et al., 2010).

Among physiological measures, EEG measures have attracted the most research attention. Traditionally, the feature extraction and electrode selection are based on neuro-scientific assumptions. Neurology and clinical research has indicated associations between emotional states and EEG powers in various frequency bands. Beta waves have been found to be associated with an active state of mind, whereas alpha waves are more dominated in a relaxed state. Therefore, prior research has used high levels of β wave power, low levels of α wave power, or large ratios of β/α to indicate high-level arousal (Choppin, 2000; Bos, 2006). Neurology findings suggests that hemispherical asymmetry can reflect emotion valence (Schmidt and Trainor, 2001). Left frontal inactivation indicates a withdrawal response and a negative emotion, whereas right frontal inactivation indicates an approach response and a positive emotion. Researchers have developed a number of measures of the asymmetry of α and β band power in the two hemispheres to indicate emotion valence (Bos, 2006; Brown et al., 2012; Davidson, 1992; Niemic and Warren, 2002). Davidson (1992) and Brown et al. (2012) measured valence using the differential asymmetry, e.g., the difference in α wave power between the left and right hemispheres of the frontal lobe. Bos (2006) measured valence using the rational asymmetry, e.g., the ratios of alpha or beta wave power between the left and right hemispheres. Furthemore, Ramirez and Vamvakousis (2012) estimated valence values by comparing the difference of α/β ratio between left and right hemispheres. In addition to this neuro-scientific approach for feature extraction and electrode selection, some researchers adopted a data-driven approach by applying computational methods (e.g., machine learning) to optimize the selection of features and electrodes from a vast amount of possible features captured by advanced signal processing technologies. Some advanced feature extraction methods, such as fractal dimension features (Liu and Sourina, 2012; Liu et al., 2011; Sourina and Liu, 2011), higher order crossings (Petrantonakis and Hadjileontiadis, 2010) and higher order spectra (Jenke et al., 2014) have been developed and found successful applications in emotion recognition.

EEG has been used to measure taste-related emotions in a limited number of studies. Park et al. (2011) used EEG to monitor negative, neutral, or positive emotions elicited by food tastes. They recognized emotion for each participant separately, using the common spatial pattern features of EEG signals from six electrodes and a support vector machine. The average accuracy of individual emotion classification was 70%. Brown et al. (2012) used the α band power at F3, F4 and Fz to measure the consumers' preference between beverages from two brands, but found no difference. Therefore, further verification is still required about whether EEG can be used to effectively measure flavor-evoked emotions.

Another family of physiological measures for emotions is that of cardiovascular measures. Emotional arousal spreads through the peripheral/autonomic nervous system. This increases the heart rate (Appelhans and Luecken, 2006). Heart rate variability—the variation in the beat-to-beat interval—has been found to be an indicator of one's ability to recognize emotions (Quintana et al., 2012). Because the peripheral nervous system plays a role in both pleasant and unpleasant emotions (Gellhorn, 1970; Kling, 1933), there is no consistent conclusion regarding the relation between HR, HRV and emotion valence. Most empirical studies using cardiovascular measures examine the relation between HR, HRV and specific emotions. Kreibig (2010) review of 134 publications summarized experimental investigations of specific emotional effects on autonomic nervous system response and found that HR was increased for both negative (anger, anxiety, contamination-related disgust, embarrassment, fear, crying sadness) and positive emotions (imagined anticipatory pleasure, happiness, joy) as well as for surprise. Increased HRV may indicate contamination-related disgust, amusement and joy, whereas decreased HRV may indicate happiness and visual anticipatory pleasure.

Emotions can also be reflected by people's facial expressions. Another measure on face is EMG. Facial expressions can be motioned by EMG, as they are caused by the contraction of facial muscles. Negative emotions make people frown, which can be reflected by the activities of the corrugator muscle, while positive emotions make people smile, which can be reflected by the activities of zygomatic muscles (Cacioppo et al., 1986; Dimberg, 1990; Lang et al., 1993; Laparra-Hernández et al., 2009; Warrenburg, 2005). Facial EMG is particularly useful in studies of emotions that are so weak that it is difficult to visually detect changes in facial expressions so that facial action coding is insensitive (Cacioppo et al., 1986). This technique has been applied to detect product-related emotions (Laparra-Hernández et al., 2009; van den Broek and Westerink, 2009).

3. Method

3.1. Participants

We recruited 24 Chinese students (14 females and 10 males) from Tsinghua University. They were aged from 19 to 27 years old ($M = 23$, $SD = 2$) with no gustatory or olfactory perception impairment, or neurological or psychiatric disorders. They were all right-handed to avoid likely individual differences of the dominant hemisphere (Provins and Cunliffe, 1972).

3.2. Stimuli

Commonly used flavors in oral products in the market fall into five categories: mint, herbal, fruit, spices, and floral flavors. First, 11 flavors from the five categories were chosen by an expert in oral care products from P & G Technologies (Beijing) Ltd. We conducted a pilot study to collect emotional responses to these flavors with 13 Chinese participants (6 females and 7 males). Based on the pilot study results, we chose stimuli for the formal experiment in the following way: one flavor in each category needed to be included and the discrimination in participants' attitudes towards stimuli had to be as large as possible. They were peppermint, wintergreen, grapefruit, clove, and rose. Each 100 g stimulus solution contained 0.08 g Tween 80, 0.016 g saccharin sodium salt, 0.5 g propylene glycol, essential oil of a flavor (0.04 g for peppermint and grapefruit, 0.06 g for wintergreen, and 0.02 g for clove and rose), and pure water.

3.3. Measures

3.3.1. Self-report measures

- **Valence, arousal, and dominance:** Three emotion dimensions in the VAD model were measured by the Self-Assessment Manikin (SAM) (Hodes et al., 1985), which was a pictorial 9-point instrument directly assessing the valence (1 for extremely pleasant and 9 for extremely unpleasant), arousal (1 for extremely aroused and 9 for extremely calm) and dominance dimensions (1 for totally controlled and 9 for totally in control). SAM is cross-cultural, concise and convenient (Bradley and Lang, 1994) and widely used in many emotion measurement situations (Liu et al., 2011; Valenza et al., 2014). In addition, the scores of emotion dimensions reported by respondents can be used to validate physiological measures.
- **Specific emotions:** six items from the EsSense Profile (King and Meiselman, 2010) were used. The original scale contained 39 items, measured by a five-point intensity scale from 1 = "not at all" to 5 = "extremely". Some items in this scale, however, were hard to understand in the oral care product context, and some items were difficult to be differentiated from each other for Chinese participants. In the pilot study, we found six items, i.e., "aggressive", "disgusted", "good", "loving", "pleasant", and "worried", could well distinguish emotions evoked by different flavors in oral care products. Therefore, we adopted these 6 items in the formal experiment.
- **Perception of flavors:** What participants think about the flavor is useful for brand or product positioning. Perception of stimuli was measured by 10 items describing features of oral care products, namely "confident", "refreshing", "premium", "natural", "familiar", "functional", "professional", "fashionable", "classic", and "unique", which were developed by two experts of oral care product designs in P & G Technologies (Beijing) Ltd. All items were measured using a five-point scale (1 = "not at all", and 5 = "extremely").
- **Overall attitude and purchase behavior intention:** Overall attitude towards a flavor was measured by a 5-point Likert scale with one item "I like this flavor". Purchase behavior intention was measured on a 5-point Likert scale with one item "I would like to purchase oral care products with this flavor".

3.3.2. EEG

In this study, we followed the neuro-scientific approach rather than the data-driven approach for the following reasons. First, EEG band power features were the most commonly used in practice due to the good expandability of the association between emotion and EEG band power features (Coan and Allen, 2004), whereas it is difficult, if not impossible, to explain the model generated by data-driven methods. Second, the data-driven methods need a training session for the classifiers. If more than two classes of valences need to be trained, which is common in product evaluation, the training sessions could become heavy and difficult to implement. Third, many of the data-driven studies used subject-dependent algorithms. But we are more interested in EEG features that are common across subjects (e.g., subject-independent). Therefore, we chose EEG spectral power features based on neuro-scientific assumptions in the current study.

We collected signals with 128 Hz sample rate in 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4) by Emotiv EPOC neuroheadset and Emotiv Testbench software. To reduce the effect of the movement of participants' tongues and muscles around the mouth on brain signals, we recorded signals after the participant had sipped the solution, closed their eyes and kept still, instead of immediately after the participant sipped the solution.

Two types of EEG features were extracted. First, we used the power spectral density (PSD, refers to the spectral energy distribution per unit time) of the β band in the parietal lobes, i.e., P7 and P8, to represent the arousal of emotion. The PSD was estimated using Welch's method. Second, we measured the valence of emotions by examining the asymmetry of the PSD in the left and right frontal lobes. Psychophysiological research has found that left frontal inactivation indicates a withdrawal response and a negative emotion, whereas right frontal inactivation indicates an approach response and a positive emotion. Because high alpha activity and low beta activity are two indicators of low brain activity, the alpha/beta ratio is a reasonable indicator of the inactivation state of mind (Bos, 2006). We compared the difference in inactivation between right and left hemisphere and used the differential value as the indicator of valence:

$$\text{positive valence} = \frac{\alpha\ \mathrm{PSD_{right}}}{\beta\ \mathrm{PSD_{right}}} - \frac{\alpha\ \mathrm{PSD_{left}}}{\beta\ \mathrm{PSD_{left}}} \tag{1}$$

where "right" and "left" denote the symmetric pairs of electrodes on the left/right hemisphere, i.e. AF4 and AF3, F4 and F3, F8 and F7, and FC6 and FC5 in our study. It should be noted that the differential value is an indicator of relative positivity of emotions. Whereas a high value indicates more positive emotions, the sign of the value (positive or negative) does not mean that the emotion is of a positive or negative valence.

3.3.3. EMG

An 8-channel biofeedback equipment (manufactured by Thought Technology Ltd) with a sampling rate of 256 Hz was used to measure participants' zygomatic EMG (in microvolts, μV). Previous research showed that increasing zygomatic muscle activity indicated positive emotions. Because all the flavors we examined were supposed to induce positive emotions, we measured only EMG of the zygomatic muscles. EMG data were collected at the moment the participant finished sipping the liquid. To mitigate the effects of individual difference, we gathered a participant's EMG under the pure water condition as the baseline of the participant. His/her EMG amplitude value under each flavor condition was "standardized" as follows:

$$\mathrm{EMG_d} = \frac{mean\, EMG\, amplitude\, in\, 10s_{flavor} - EMG\, amplitude_{baseline}}{EMG\, amplitude_{baseline}} \tag{2}$$

3.3.4. HR and HRV

HR and HRV data were also collected using the 8-channel biofeedback equipment. Blood volume pulse (BVP) was gathered at a sampling rate of 256 Hz. The electrode was placed on the participant's thumb of the left hand. BVP data were collected at the moment the participant finishing sipping the liquid. BVP data were further transformed into HR and HRV. HR was the mean of reciprocals of time intervals between each two peaks of BVP. HRV was the standard deviation of reciprocals of time intervals between each two peaks of BVP. To mitigate the effects of individual difference, each participant's HR and HRV under pure water conditions were regarded as baselines of the participant. His/her HR and HRV values under each flavor condition were "standardized" as follows:

$$\mathrm{HR_d} = \frac{HR_{flavor} - HR_{baseline}}{HR_{baseline}} \tag{3}$$

$$\mathrm{HRV_d} = \frac{HRV_{flavor} - HRV_{baseline}}{HRV_{baseline}} \tag{4}$$

3.4. Procedure

To avoid the interference between head-mounted devices and also to avoid intrusive feelings from the participants, we separated participants into an EEG group and an EMG group. Twelve participants in the EEG group were measured by questionnaire (self-report measures), EEG, HR and HRV, whereas the 12 participants in the EMG group were measured by questionnaire, EMG, HR and HRV.

The experiment for each participant lasted about 1 h. Each participant experienced six conditions: five flavors and the pure water condition as the baseline. Each condition was replicated three times for each participant, i.e., 18 trials ((5 flavors + pure water) * 3) for each participant. The order of trials was counterbalanced to ensure that (1) different flavors were separated for each participant, (2) that the three trials of a certain flavor had different previous flavors for each participant, and (3) the order of flavors were counterbalanced across participants. In each trial, the participant sat in a comfortable chair and sipped the stimuli using a straw. The participants in the EMG group (1) rinsed their mouth with pure water, (2) sipped 20 ml solution and kept in the mouth for 10 s (at the moment the participant finished taking in the liquid, we started to record EMG and BVP) and (3) spat the solution out and filled out the questionnaire. For the EEG measurement, to reduce the effect of the movement of their tongues, muscles around the mouth, and eye movements on brain signals, we recorded signals after the participant had sat still, instead of immediately when the participant sipped the solution. The participants in the EEG group (1) rinsed their mouth with pure water, (2) sipped 20 ml solution and kept it in the mouth for 10 s (at the moment the participant finished taking in the liquid, we started to record BVP; when he/she had closed eyes, we started to record EEG signals for 10 s) and (3) spat the solution out and filled in the questionnaire (see Fig. 1).

4. Results

4.1. Sensitivity – Distinguish emotions evoked by different flavors

4.1.1. SAM

The sensitivities of different measurements were compared by examining how self-report, EEG, EMG, HR and HRV could distinguish emotions evoked by different flavors. First, the data of SAM violated the normality assumptions for parametric analysis, therefore we conducted Friedman tests and Wilcoxon signed rank tests for post hoc tests. The Friedman tests showed significant differences among the five flavors in arousal in SAM (1–9 indicated aroused to calm) and valence in SAM (1–9 indicated positive to negative) (p values < .05, see Table 1). The "calm" score in SAM of grapefruit was higher than the score of rose, peppermint and wintergreen, and the calm scores of rose, peppermint, and wintergreen were higher than that of clove (p values < .05). The "negative" score of clove in SAM was higher than the scores of wintergreen and rose, and the negative scores of wintergreen and

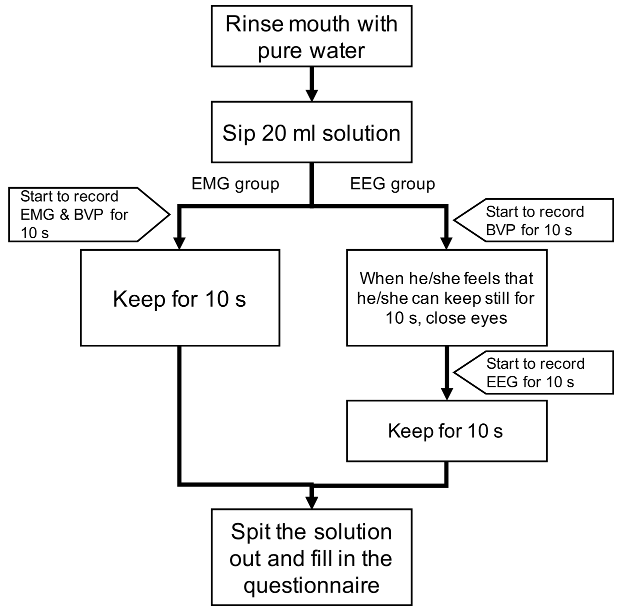

Fig. 1. Procedure of each trial.

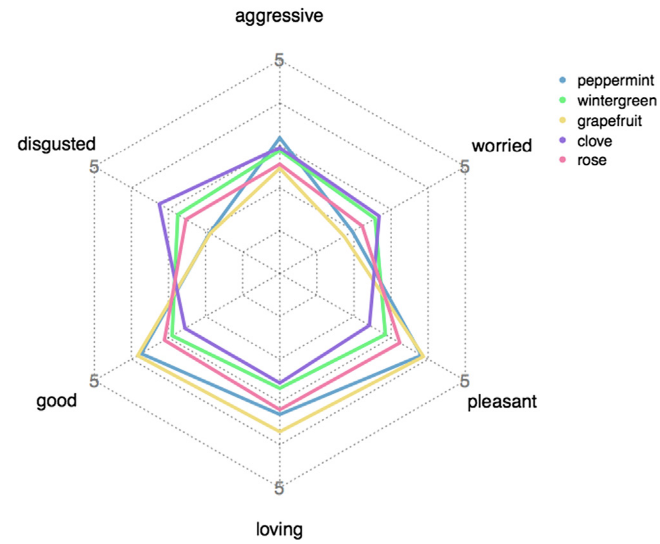

Fig. 2. The score of the six items in the EsSense Profile for specific emotions of the five flavors.

rose were higher than those of peppermint and grapefruit (p values < .05).

4.1.2. Specific emotions

Friedman tests showed significant differences among the five flavors in six items from the EsSense Profile for specific emotions (p values < .01, see Fig. 2 for mean values). Wilcoxon signed rank tests showed that the "aggressive" score of peppermint and cloves was higher than the scores of grapefruit and rose, and the "aggressive" score of wintergreen was higher than that of grapefruit. The "disgusted" scores of grapefruit and peppermint were lower than those of other flavors. The "good" and "pleasant" scores of grapefruit and peppermint were higher than those of other flavors. The "loving" scores of grapefruit, peppermint and rose were higher than those of wintergreen and clove. The "worried" scores of grapefruit and peppermint were lower than those of wintergreen and clove.

4.1.3. Perception of flavors, overall attitude and purchase intention

Friedman tests showed significant differences among the five flavors in overall attitude (1–5 indicated dislike to like), purchase intention, and 10 items of perception of flavors (p < .01). Wilcoxon signed rank tests showed that the "like" (attitude) scores of peppermint and grapefruit were significantly higher than that of rose (p < .05), the attitude score of rose was significantly higher than that of wintergreen (p < .05), and the attitude score of wintergreen was marginally significantly higher than that of clove (p < .1). The "purchase" scores of peppermint and grapefruit were higher than those of rose and wintergreen (p < .05), and the purchase scores of rose and wintergreen were higher than that of clove (p < .05).

4.1.4. EEG

Fig. 3 shows the difference of α/β PSD ratios at AF4 and AF3 (AF4-AF3 ratio in the following text) and the P8 β PSD values of the

Table 1
Mean values, standard deviations, and results of Friedman tests for questionnaire items.

	Peppermint	Wintergreen	Grapefruit	Clove	Rose	χ^2_4	P value
Arousal in SAM (Calm)	4.90 (2.20)	4.71 (2.21)	5.82 (1.67)	4.12 (2.11)	5.15 (2.00)	<.01	21.39
Valence in SAM (Negative)	3.47 (1.84)	5.21 (2.13)	3.46 (1.72)	5.93 (2.18)	4.62 (2.11)	<.01	52.01
Dominance in SAM	5.18 (1.76)	5.11 (1.85)	4.64 (1.29)	5.57 (2.02)	4.96 (1.54)	.35	4.4
Aggressive	3.18 (1.15)	2.88 (1.11)	2.46 (0.99)	2.94 (1.07)	2.56 (1.05)	<.01	22.63
Disgusted	1.89 (1.01)	2.75 (1.26)	1.88 (0.96)	3.25 (1.36)	2.53 (1.29)	<.01	41.36
Good	3.74 (0.90)	2.90 (0.89)	3.83 (0.77)	2.56 (1.12)	3.11 (1.07)	<.01	43.59
Loving	3.29 (1.07)	2.68 (1.03)	3.69 (0.93)	2.56 (1.10)	3.18 (1.14)	<.01	32.90
Pleasant	3.82 (1.09)	2.85 (1.03)	3.88 (0.87)	2.42 (1.17)	3.24 (1.17)	<.01	41.96
Worried	1.96 (0.91)	2.57 (1.11)	1.74 (0.71)	2.68 (1.11)	2.22 (1.01)	<.01	41.62
Confident	3.61 (0.85)	2.97 (0.86)	3.36 (0.79)	2.60 (0.88)	2.97 (0.92)	<.01	41.44
Refreshing	4.56 (0.53)	3.19 (1.00)	3.42 (0.88)	3.07 (0.98)	3.01 (0.94)	<.01	54.05
Premium	3.31 (0.87)	2.75 (0.90)	3.15 (0.94)	2.50 (0.87)	2.82 (1.04)	<.01	23.59
Natural	3.24 (1.04)	2.39 (1.03)	3.6 0 (1.11)	2.29 (1.07)	2.92 (1.20)	<.01	31.37
Familiar	3.72 (1.08)	2.71 (1.05)	3.68 (1.10)	2.31 (1.22)	3.15 (1.11)	<.01	35.42
Functional	4.01 (0.72)	3.11 (1.00)	3.26 (0.86)	2.94 (0.87)	2.99 (0.81)	<.01	45.52
Professional	3.71 (0.91)	2.99 (1.03)	3.12 (0.85)	3.00 (0.93)	2.82 (0.84)	<.01	23.77
Fashionable	2.97 (0.87)	2.53 (0.90)	3.19 (1.10)	2.42 (0.93)	2.83 (1.16)	<.01	17.42
Classic	3.88 (0.99)	2.61 (1.07)	2.97 (0.89)	2.29 (0.96)	2.62 (0.80)	<.01	51.88
Unique	2.71 (0.96)	3.26 (1.06)	2.96 (1.00)	3.40 (1.21)	3.12 (1.19)	0.03	10.51
Overall attitude (Like)	3.67 (1.06)	2.44 (1.02)	3.75 (1.02)	2.06 (1.01)	3.11 (1.09)	<.01	51.33
Purchase intention	3.71 (1.18)	2.43 (1.10)	3.58 (1.12)	2.01 (1.07)	2.78 (1.22)	<.01	52.12

Note: arousal, valence, and dominance are scales from 1 to 9, whereas other items are from 1 to 5.

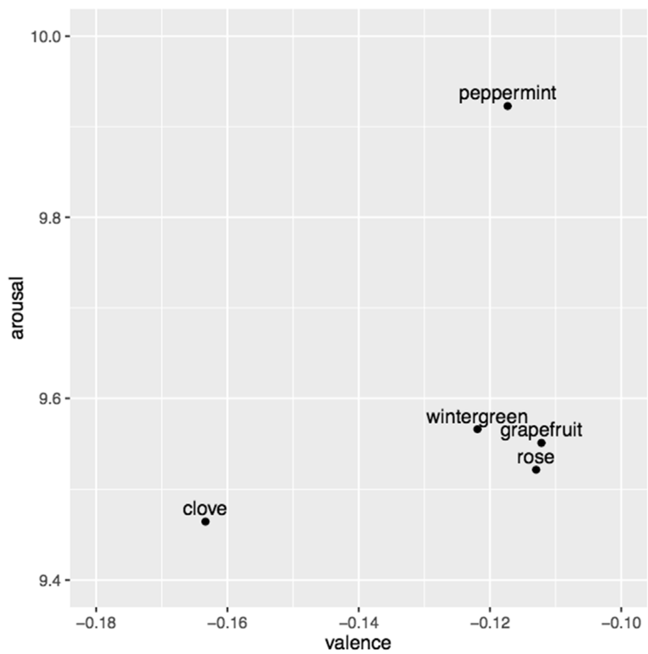

Fig. 3. The mean values of the EEG features for emotion valence (AF4-AF3 ratio) and arousal (P8 β PSD) for the five flavors.

five flavors. To compare the difference among EEG features under the five flavor conditions, we conducted repeated measures ANOVA, and found no significant difference among the AF4-AF3 ratio of five flavors ($F_{4,44} = 1.64$, p value = .18, generalized $\eta^2 = .02$). Further, pairwise T tests with no correction showed that the AF4-AF3 ratio of clove (M = −0.14, SD = 0.18) was lower than that of grapefruit (M = −0.12, SD = 0.20, p = .06 marginally significant), rose (M = −0.10, SD = 0.19, p < .01), peppermint (M = −0.12, SD = 0.20, p = .08), and wintergreen (M = −0.12, SD = 0.02, p = .03) (see Fig. 3). No significant difference was found in power spectrum densities in parietal lobes under the different flavor conditions.

4.1.5. EMG

The EMG data of two participants were dismissed due to the poor contact of the electrode. Repeated measures ANOVA was conducted to study the effect of flavor on standardized EMG values on the zygomatic muscles (see Table 2). We found no significant difference of standardized EMG values when participants tasted peppermint, wintergreen, grapefruit, clove and rose. The flavor did not significantly affect EMG values ($F_{1.81, 16.31} = 1.31$, p = .28, generalized $\eta^2 = .07$, the degree of freedom was corrected by Greenhouse–Geisser epsilon).

4.1.6. HR and HRV

Repeated measures ANOVA was conducted to study the effect of flavor on standardized HR and HRV values. We found no significant difference of standardized HR values when participants tasted peppermint, wintergreen, grapefruit, clove and rose ($F_{4,92} = 0.37$, p = .83, generalized $\eta^2 < .01$).

The effect of flavor was marginally significant on standardized HRV values ($F_{4,92} = 2.09$, p = .09, generalized $\eta^2 < .01$). Post hoc analysis, i.e., pairwise paired *t*-test with no correction, showed that standardized HRV value of clove (M = 0.12, SD = 0.61) was significantly lower than that of wintergreen (M = 0.26, SD = 0.60, p = .03), grapefruit (M = 0.23, SD = 0.62, p = .05) and marginally significantly lower than rose (M = 0.23, SD = 0.74, p = 0.09). No significant difference was found between standardized HRV under the condition of peppermint (M = 0.21, SD = 0.63) and standardized HRV under the condition of the other four flavors.

4.2. Reliability

Participants tasted each flavor three times in this study. Reliability was compared by examining whether the results from the three measures for each flavor and participant were consistent. The consistency of the three measures of the same flavor and the same participant was calculated in terms of Cronbach's α coefficient. Cronbach's α is usually used to indicate the consistency among several items that are supposed to describe one common variable. Here we used Cronbach's α to indicate the average degree of interrelatedness among the three measures using the same method (see Table 3). The EEG feature, AF4-AF3 ratio, had the highest reliability (Cronbach's $\alpha = .85$). Cronbach's α coefficients of HR, HRV, valence in the SAM, selected EsSense Profile items, perceptions, attitude and purchase intention were higher than .70, and their reliability was acceptable. However, the arousal dimension of the SAM and the EMG measure did not reach the general accepted level of reliability (<.70).

4.3. Correlation among measures

We assessed convergent validity and divergent validity of different measurements by analyzing the correlations among them. Convergent validity refers to the agreement between instruments that are expected to measure the same or similar constructs, whereas divergent validity tests whether measurements that are not supposed to be related are in fact unrelated (Campbell and Fiske, 1959). The result of Spearman correlation analysis among the measures is shown in Table 4.

First, we examined the convergent validity of measures. The β PSD of parietal lobes, HR, and the calm dimension of SAM were supposed to measure emotional arousal. The Spearman correlation showed significant correlation between P8 β PSD and calm in SAM (ρ = - .19, p = .009), between P8 β PSD and HR ($\rho = .22$, p = .003), and between HR and calm in SAM ($\rho = -.20$, p < .001). No significant result was found for P7 β PSD of the parietal lobes. In addition, HR also positively correlated with the high arousal item, aggressive ($\rho = .12$, p = .02). The convergent validity of these arousal measures were established.

The AF4-AF3 ratio, HRV, and Zygomatic EMG was supposed to reflect valence of emotion. The correlation analysis showed that the AF4-AF3 ratio was marginally significantly correlated with positive valence in SAM ($\rho = -.12$, p = .10) and overall attitude ($\rho = .14$, p = .07). The AF4-AF3 ratio was positively correlated with two

Table 2
Mean values, standard deviations and results of repeated measures ANOVA for standardized EMG, HR, and HRV.

	Peppermint	Wintergreen	Grapefruit	Clove	Rose	F	P value
EMG	−0.01 (0.25)	−0.08 (0.35)	−0.06 (0.23)	0.12 (0.48)	−0.01 (0.28)	1.31	0.28
HR	0.01 (0.09)	−0.004 (0.09)	−0.003 (0.08)	−0.002 (0.09)	−0.01 (0.09)	0.37	0.83
HRV	0.21 (0.63)	0.26 (0.60)	0.23 (0.62)	0.12 (0.61)	0.23 (0.74)	2.09	0.09

Note: EMG, HR, and HRV were standardized by each participant's baseline.

Table 3
Cronbach's α coefficients of measures.

Measures	α	Measures	α	Measures	α
Selected EsSense Profile items		**Perceptions**		**Attitude**	0.83
aggressive	0.77	confident	0.79	**Purchase intention**	0.81
disgusted	0.77	refreshing	0.79	**EEG**	
good	0.82	premium	0.72	α-β ratio of AF3 and AF4	0.85
loving	0.81	natural	0.79	β power at P8	0.77
pleasant	0.82	familiar	0.80	**EMG**	0.63
worried	0.80	functional	0.70	**HR**	0.72
SAM		professional	0.77	**HRV**	0.81
arousal (calm)	0.66	fashionable	0.79		
valence (negative)	0.81	classic	0.76		
dominance	0.70	unique	0.77		

Table 4
Correlation between EEG, EMG, HR, HRV, and questionnaire items.

	Arousal in SAM (Calm)		Valance in SAM (Negative)		AF4-AF3 ratio		P8 β		EMG		HR		HRV	
	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P
Overall attitude (like)	0.40	<0.001	−0.72	<0.001	0.14	0.07	−0.08	0.31	−0.08	0.32	−0.14	0.01	0.09	0.08
Purchase intention	0.42	<0.001	−0.63	<0.001	0.12	0.11	0.10	0.17	−0.13	0.12	−0.13	0.02	0.11	0.04
Aggressive	−0.46	<0.001	0.20	<0.001	−0.17	0.02	0.06	0.43	−0.03	0.70	0.12	0.02	0.03	0.59
Disgusted	−0.35	<0.001	0.66	<0.001	−0.07	0.39	−0.001	0.98	0.12	0.15	0.09	0.101	−0.05	0.38
Good	0.46	<0.001	−0.79	<0.001	0.16	0.04	−0.23	0.002	−0.13	0.12	−0.10	0.06	−0.01	0.90
Loving	0.41	<0.001	−0.76	<0.001	0.13	0.07	−0.29	<0.001	−0.06	0.51	−0.09	0.101	−0.06	0.27
Pleasant	0.45	<0.001	−0.79	<0.001	0.11	0.16	−0.19	0.01	−0.09	0.29	−0.13	0.01	0.02	0.76
Worried	−0.29	<0.001	0.60	<0.001	0.05	0.47	0.01	0.88	0.11	0.19	0.04	0.44	0.09	0.096
Confident	0.04	0.50	−0.54	<0.001	0.05	0.52	−0.04	0.61	−0.02	0.83	0.003	0.95	−0.06	0.26
Refreshing	−0.08	0.13	−0.22	<0.001	−0.02	0.84	0.15	0.04	−0.11	0.20	−0.02	0.68	0.05	0.38
Premium	0.12	0.02	−0.38	<0.001	0.12	0.11	0.06	0.39	−0.10	0.27	−0.15	0.004	0.16	0.002
Natural	0.46	<0.001	−0.62	<0.001	0.11	0.14	0.13	0.08	−0.17	0.04	−0.12	0.03	0.07	0.21
Familiar	0.36	<0.001	−0.61	<0.001	0.09	0.22	0.03	0.67	−0.05	0.51	−0.08	0.13	0.01	0.93
Functional	−0.07	0.21	−0.29	<0.001	−0.10	0.18	0.22	0.002	0.10	0.21	0.04	0.46	−0.03	0.56
Professional	−0.06	0.26	−0.16	0.003	−0.09	0.25	0.15	0.04	0.17	0.04	0.08	0.15	0.01	0.86
Fashionable	0.07	0.16	−0.39	<0.001	−0.08	0.27	0.10	0.17	0.13	0.12	−0.10	0.049	0.06	0.26
Classic	0.20	<0.001	−0.38	<0.001	0.05	0.55	0.15	0.04	−0.01	0.88	−0.11	0.045	0.10	0.06
Unique	−0.20	<0.001	0.21	<0.001	0.10	0.20	0.00	1.00	0.01	0.88	−0.10	0.054	−0.01	0.84
Arousal in SAM (Calm)	–	–	−0.43	<0.001	0.25	<0.001	−0.19	0.01	−0.16	0.06	−0.20	<0.001	0.13	0.01
Valance in SAM (Negative)			–	–	−0.12	0.097	0.30	<0.001	0.06	0.49	0.05	0.32	0.09	0.098
AF4-AF3 ratio					–	–	−0.30	0.60	–	–	−0.03	0.66	−0.01	0.86
P8 β							–	–	–	–	0.22	0.003	0.005	0.95
EMG									–	–	0.15	0.08	−0.15	0.08
HR											–	–	−0.53	<0.001
HRV													–	–

positive EsSense Profile terms, i.e., good ($\rho = .16$, $p = .04$), loving ($\rho = .13$, $p = .07$), and negatively correlated with one negative term, aggressive ($\rho = -.17$, $p = .02$). HRV and zygomatic EMG, however, did not correlate with the valence in SAM, specific items in the EsSense Profile, and the AF4-AF3 ratio (the EEG feature for valence). In addition, zygomatic EMG was marginally significantly correlated with calm in SAM ($\rho = -.16$, $p = .06$) and high HR ($\rho = .15$, $p = .08$). This suggested that zygomatic EMG may indicate emotion arousal evoked by flavors, instead of emotional valence. The results provide evidence for the convergent validity of AF4-AF3 ratio and the calm dimension of SAM measures, but not for HRV and Zygomatic EMG.

Second, we examined the divergent validity of arousal and valence measures. Arousal and valence are supposed to be orthogonal dimensions. Therefore, the measurements for these two dimensions should be uncorrelated. The results showed that whereas the EEG measures of arousal (AF4-AF3 ratio) and valence (P8 β PSD) measures were uncorrelated. There was, however, a medium to high correlation between the arousal and the valence dimensions of SAM ($\rho = -.43$, $p < .001$). Furthermore, the AF4-AF3 ratio, which was supposed to measure valence, was also significantly correlated with calm in SAM ($\rho = .25$, $p < .001$). The result supported the divergent validity of EEG measures, whereas the divergent validity of SAM was inadequate.

4.4. Predicting overall attitude and purchase intention

4.4.1. Self-report measures, EEG, HR and HRV

To exclude the effect from individual difference, we nested the data into different participants and conducted multilevel linear regression. The dependent variables were the overall attitude to a flavor and the purchase intention. The independent variables were the scores of self-report measures (six EsSense Profile items, perception of flavors), the EEG feature (AF4-AF3 ratio), and the standardized EMG, HR, and HRV. To ascertain whether there is variation over different individuals, we fitted a baseline model (I) in which we included only the intercept. Next, we fitted a model (RI) that allowed intercepts to vary over individuals. The ANOVA indicated that for predicting the overall attitude, model RI yielded significantly lower AIC (Akaike information criterion, 1163.83) and BIC (Bayesian information criterion, 1175.49) than model I did (AIC = 1174.25, BIC = 1182.03, $\chi^2_{(1)} = 12.42$, $p < .001$); for predicting the purchase intention, model RI yielded lower AIC (1194.00) and BIC (1205.65) than model I did (AIC = 1218.77, BIC = 1226.54, $\chi^2_{(1)} = 26.78$, $p < .001$). This suggested the necessity of multilevel linear regression.

The data of the 12 participants in the EEG group were analyzed. To predict overall attitude towards a flavor, self-report measures (six EsSense Profile items, perception of flavors), the EEG feature (AF4-AF3 ratio), HR, and HRV were entered as fixed effects to fit a model RI_EEG_A. The likelihood ratio test suggested that the model was significantly improved ($\chi^2_{(19)} = 240.16$, $p = < .001$). Model RI_EEG_A for predicting overall attitude yielded lower AIC (374.19) and BIC (444.43) than model RI did (AIC = 576.35, BIC = 585.92. In the model RI_EEG_A (see Table 5), aggressive, disgusted, loving, refreshing, fashionable, classic, and AF4-AF3 ratio were significant predictors.

To predict purchase intention, the independent variables above were also entered in the RI model as fixed effects to fit a model RI_EEG_P. The likelihood ratio test suggested that the model was significantly improved ($\chi^2_{(19)} = 190.28$, $p = <.001$). Model RI_EEG_P yielded lower AIC (412.65) and BIC (482.89) than model RI (AIC = 564.93, BIC = 574.51). In the model RI_EEG_P, confident, premium, natural, fashionable and classic were significant predictors. None of the physiological measures was a significant predictor.

4.4.2. Self-report, EMG, HR, and HRV

The data of the 12 participants in the EMG group were analyzed. To predict overall attitude towards a flavor, self-report measures (six EsSense Profile items, perception of flavors), EMG, HR and HRV were entered as fixed effects to fit a model RI_EMG_A. The likelihood ratio test suggested that the model was significantly improved ($\chi^2_{(19)} = 219.29$, $p = < .001$). Model RI_EMG_A yielded lower AIC (295.41) and BIC (361.49) than model RI (AIC = 476.70, BIC = 485.71). In the model RI_EMG_A (see Table 6), good, pleasant and fashionable were the significant predictors. None of the physiological measures was a significant predictor.

To predict purchase intention, the independent variables above were also entered in the RI model as fixed effects to fit a model RI_EMG_P. The likelihood ratio test suggested that the model was significantly improved ($\chi^2_{(19)} = 229.46$, $p = <.001$). Model RI_EMG_P yielded lower AIC (318.11), and BIC (384.19) than model RI (AIC = 509.56, BIC = 518.57). In the model RI_EMG_P, good, pleasant, and natural were the significant predictors. None of the physiological measures was a significant predictor.

5. Discussion

SAM has been a widely used instrument for measuring emotional responses to various stimuli. Our study shows that the arousal and the valence dimensions are able to distinguish emotions evoked by the five flavors ($p < .05$ for Friedman tests). The reliability of SAM in repetitive tests, however, is not satisfactory. Only the dimension of valence reaches a good level of reliability (Cronbach's α: 0.81), whereas the arousal dimension (Cronbach's α: 0.66) and the dominance dimension (Cronbach's α: 0.70) are questionable or barely acceptable. Furthermore, we found that the validity of SAM results is questionable, at least for the Chinese participants in our study. The SAM shows only pictures without textual explanation. This leaves room for individuals' interpretation. We interviewed five participants for their understanding of SAM. Most of them did not understand the meaning of the dominance pictures. Some of them thought the arousal pictures represented the extent of anger. The correlation analysis suggests that arousal in the SAM significantly correlates with valence in SAM and the EEG feature AF4-AF3 α-β ratio. This indicates a significant association between the valence and the arousal dimensions, whereas these two are theoretically orthogonal dimensions. The less satisfactory reliability and divergent validity, and the problems in interpreting the pictures suggest that SAM may not be an ideal instrument for measuring flavor-evoked emotions of Chinese consumers.

The selected EsSense Profile items and the perceptions of flavors provides detailed and easy-to-interpret information which are hard to obtain with physiological measurements. These measures are sensitive enough to distinguish emotions evoked by the five flavors ($p < .05$ for Friedman tests of all items except "unique"). The selected EsSense Profile items have acceptable reliability, as indicated by Cronbach's α coefficients ranging from .70 to .82. These items are also the most important predictors of overall attitude and purchase intention, as indicated by the results from both the EMG and the EEG group. These items, however, may be influenced by the halo effect (Murphy et al., 1993; Thorndike, 1920), i.e., that the overall impressions of the flavor strongly influence the ratings of specific perceptions of the flavor. For example, clove oil is generally used in dentistry to treat pain, and its scent often lingers in the dental clinics. Therefore, clove flavor should be associated with the

Table 5
Multilevel linear models for overall attitude and purchase intention (independent variables were questionnaire factors and the EEG feature AF4-AF3 ratio, HR, and HRV).

	Overall attitude				Purchase intention			
	Value	Standard error	t_{149}	P value	Value	Standard error	t_{149}	P value
(Intercept)	0.058	0.464	0.126	0.9	−0.106	0.616	−0.172	0.863
aggressive	0.18	0.065	2.77	0.006	0.011	0.074	0.145	0.885
disgusted	−0.175	0.066	−2.651	0.009	−0.089	0.072	−1.232	0.22
good	−0.017	0.108	−0.159	0.874	0.058	0.117	0.496	0.621
loving	0.379	0.101	3.761	<.001	0.095	0.114	0.837	0.404
pleasant	0.187	0.107	1.748	0.082	−0.063	0.115	−0.552	0.582
worried	−0.023	0.087	−0.261	0.794	−0.061	0.106	−0.572	0.568
confident	−0.082	0.094	−0.878	0.381	0.216	0.108	2.006	0.047
refreshing	0.165	0.074	2.221	0.028	0.063	0.081	0.774	0.44
premium	−0.011	0.081	−0.13	0.896	0.185	0.089	2.078	0.039
natural	0.092	0.07	1.316	0.19	0.168	0.078	2.137	0.034
familiar	−0.044	0.079	−0.558	0.577	0.061	0.087	0.704	0.483
functional	−0.093	0.085	−1.1	0.273	0.02	0.093	0.215	0.83
professional	0.06	0.08	0.749	0.455	0.019	0.089	0.21	0.834
fashionable	0.245	0.073	3.349	0.001	0.177	0.08	2.208	0.029
classic	0.219	0.076	2.888	0.004	0.208	0.083	2.511	0.013
unique	−0.064	0.06	−1.058	0.292	−0.033	0.069	−0.477	0.634
AF4-AF3 ratio	1.085	0.429	2.528	0.013	0.463	0.498	0.93	0.354
HR	−1.434	0.917	−1.563	0.12	−1.293	1.021	−1.266	0.208
HRV	−0.039	0.126	−0.312	0.755	0.157	0.139	1.135	0.258

Table 6
Multilevel linear models for overall attitude and purchase intention (independent variables were questionnaire factors, EMG, HR and HRV).

	Overall attitude				Purchase intention			
	Value	Standard error	t_{120}	P value	Value	Standard error	t_{120}	P value
(Intercept)	0.83	0.558	1.486	0.14	−0.477	0.616	−0.775	0.44
aggressive	−0.05	0.067	−0.74	0.461	−0.021	0.073	−0.295	0.769
disgusted	−0.064	0.077	−0.822	0.413	−0.029	0.083	−0.352	0.725
good	0.271	0.099	2.728	0.007	0.308	0.107	2.889	0.005
loving	0.036	0.095	0.376	0.708	−0.197	0.103	−1.907	0.059
pleasant	0.299	0.088	3.39	0.001	0.422	0.094	4.476	<.001
worried	−0.084	0.08	−1.055	0.293	−0.142	0.086	−1.64	0.104
confident	0.005	0.081	0.064	0.949	0.037	0.087	0.418	0.677
refreshing	0.025	0.063	0.393	0.695	0.131	0.068	1.938	0.055
premium	0.033	0.081	0.413	0.681	0.052	0.087	0.599	0.55
natural	0.09	0.073	1.232	0.22	0.2	0.078	2.557	0.012
familiar	0.099	0.069	1.419	0.158	0.128	0.075	1.71	0.09
functional	0.031	0.089	0.347	0.729	0.134	0.095	1.404	0.163
professional	−0.067	0.081	−0.833	0.406	−0.109	0.086	−1.258	0.211
fashionable	0.146	0.077	1.899	0.06	0.01	0.083	0.117	0.907
classic	−0.031	0.065	−0.485	0.628	0.029	0.07	0.412	0.681
unique	−0.071	0.059	−1.219	0.225	0.054	0.064	0.851	0.397
EMG	−0.176	0.169	−1.041	0.3	−0.202	0.183	−1.103	0.272
HR	0.237	0.8	0.297	0.767	0.091	0.855	0.107	0.915
HRV	−0.043	0.129	−0.334	0.739	0.065	0.141	0.46	0.646

perception of “functional” for many consumers, though the flavor may not be pleasant. In our study, however, clove was reported as the least functional flavor. We found that this may be due to the overall negative impressions of clove. Both SAM and the AF4-AF3 ratio indicate that clove evokes the most negative emotions. The correlation analysis also shows that the item “functional” correlates with valence in the SAM ($\rho = -.29$, $p < .001$). This indicates the occurrence of halo effect, i.e., the perception of negative valence of clove influence the rating of functional perception.

In EEG features, the AF4-AF3 α-β ratio is sensitive to distinguishing clove from others. The Cronbach's α of the AF4-AF3 α-β ratio in three trials is .85, which is the highest among all measures in this study. The AF4-AF3 α-β ratio correlates with positive emotions, i.e., marginally significantly correlates with overall attitude, positive valence in SAM, and good and loving in EsSense Profile. P8 β PSD positively correlated with the arousal dimension in SAM and HR, which is supposed to indicate emotion arousal level. These correlations are consistent with previous studies (Bos, 2006; Brown et al., 2012; Choppin, 2000; Davidson and Tomarken, 1989; Niemic and Warren, 2002). The AF4-AF3 α-β ratio is also a significant predictor of the participants' overall attitude towards a flavor. These results show that the AF4-AF3 α-β ratio and P8 β PSD can be used as indicators of flavor-evoked emotional valence and arousal level, respectively. In addition, AF4-AF3 α-β ratio and P8 β PSD were not correlated, indicating adequate divergent validity.

Although HR is not sensitive enough to distinguish the five flavors, it can indicate emotion arousal with an acceptable reliability (Cronbach's α coefficient = .72). Consistent with previous research (Appelhans and Luecken, 2006), high HR is related with emotions with high arousal level. It positively correlated with the arousal dimension in SAM and P8 β PSD. HRV can distinguish emotions evoked by the clove flavor from emotions evoked by the wintergreen, grapefruit and rose flavors, with a Cronbach's α coefficient of .81. However, HRV does not correlate with emotion valences, and both the correlation of HRV and self-report and the correlation of the EEG feature AF4-AF3 α-β ratio are not significant. Previous research also cannot provide a consistent conclusion regarding the relation between HRV and emotion valence (Kreibig, 2010), and there is a lack of physiological evidences or clinical research about the relationships of HRV and specific emotions in the EsSense Profile. In addition, some researchers were concerned that HRV effects were less consistent and often nonsignificant in taste-related measures, because reliable HR measurements require relatively long periods of time, whereas food evoked emotions in short time periods (de Wijk et al., 2012). Therefore, it is hard to use HRV to indicate emotion valence.

EMG of the zygomatic muscles is neither sensitive (cannot distinguish emotions evoked by the five flavors) nor reliable (the Cronbach's α: .63). Previous researchers found that positive emotions make people smile and increase the activities of the zygomatic and levator muscles (Cacioppo et al., 1986; Dimberg, 1990; Lang et al., 1993; Warrenburg, 2005). However, in the present study, the EMG level at the zygomatic muscles does not correlate with emotion valence. Instead, it correlates with high emotion arousal level in SAM (marginally significant) and high HR (which is supposed to indicate emotion arousal level). The possible reason may be that flavors with high arousal directly stimulate the muscles near the mouth, therefore increase zygomatic EMG. In other words, high zygomatic EMG may result from both smiles (positive valence) and stimulation from flavors (high arousal), and the effects from them are hard to separate. Therefore, it is hard for zygomatic EMG to indicate the flavor-evoked emotions reliably and validly.

6. Conclusion

To the best of our knowledge, no study has been conducted to compare different measures of flavor-evoked emotions. Our study provides a comprehensive picture of state-of-the-art instruments for recognizing flavor-evoked emotions based on empirical evidence. This picture helps designers to select proper measurements based a clear understanding of their strengths and limitations.

The results showed that self-reporting measures can distinguish different flavor products, but the inadequate divergent validity indicated that the people's explicit expression of emotions may be biased by halo effect, e.g., their overall impression of the flavor influencing their perception about specific properties of emotions (e.g., arousal and valence). The participants' perception of arousal is influenced by their perception of valence. These results suggested that self-reporting measures are good predictors to expressed overall preference and purchase intention with a flavor, but they provide less accurate information to understand specific emotional dimensions.

The two EEG measures adopted in the current study have been found valid and reliable measurement of arousal and valence of emotions, and they exhibited adequate divergent validity, i.e., the measures reflecting the orthogonal dimensions of emotions are uncorrelated. The measures are extracted based on sound neuroscientific assumptions and our results validate the capacity of these measures for measuring flavor-evoked emotions.

HR has been found a valid and reliable measure of arousal. Given the low cost and high convenience to obtain HR data, as compared with EEG measures, we found it a good supplement psychophysiological measure to the self-reporting measures, with which the measurement of arousal is often contaminated with one's perception of valence.

Finally, we did not find adequate evidence to support the validity of HRV and EMG of zygomatic for measuring flavor-evoked emotions.

This study has some limitations. Because the participants were split into EEG group and EMG group due to the difficulty to wear the two equipment simultaneously, the sample size of the EEG tests was only half of the sample size of the tests of self-reporting measures. A smaller sample size means a lower level of power for a statistic test to reject the null hypotheses. This may contribute to the result that EEG measures did not distinguish flavors as much as self-reporting measures did. To draw a fair conclusion about the relative sensitivity of different measures, a study involving equal sample sizes for different is needed. In addition, this study focused on commonly used spectral power features with sound neuropsychological assumptions. These measures, however, are likely to be subject to the influences of noises in physiological signals. Future research may explore more advanced feature extraction methods, such as fractal dimension, higher order crossings and higher order spectra (Jenke et al., 2014), and data-driven methods (Liu et al., 2011; Sourina and Liu, 2011) for recognize emotions without the constraints of physiological or clinical research support.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Project no. 71401087 and 71671102).

References

Appelhans, B.M., Luecken, L.J., 2006. Heart rate variability as an index of regulated emotional responding. Rev. Gen. Psychol. 10 (3), 229.

Berridge, K., Winkielman, P., 2003. What is an unconscious emotion? (The case for unconscious "liking"). Cognit. Emot. 17 (2), 181–211.

Bos, D.O., 2006. EEG-based emotion recognition. The influence of visual and auditory stimuli, pp. 1–17.

Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatr. 25 (1), 49–59.

Brown, C., Randolph, A.B., Burkhalter, J.N., 2012. The story of taste: using EEGs and self-reports to understand consumer choice. Kennesaw J. Undergrad. Res 2 (1), 5.

Cacioppo, J.T., Petty, R.E., Losch, M.E., Kim, H.S., 1986. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. J. Pers. Soc. Psychol. 50 (2), 260.

Campbell, D.T., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull. 56 (2), 81.

Choppin, A., 2000. EEG-based human interface for disabled individuals: Emotion expression with neural networks (Unpublished master's thesis). Tokyo Institute of Technology, (Yokohama, Japan).

Coan, J.A., Allen, J.J., 2004. Frontal EEG asymmetry as a moderator and mediator of emotion. Biol. Psychol. 67 (1), 7–50.

Czerwinski, M., Horvitz, E., Cutrell, E., 2001. Subjective duration assessment: an implicit probe for software usability. In: Proceedings of IHM-HCI 2001 conference, Vol. 2, pp. 167–170.

Damasio, A., 2006. Descartes' Error: Emotion, Reason and the Human Brain. Random House, London.

Davidson, R.J., 1992. Anterior cerebral asymmetry and the nature of emotion. Brain Cognit. 20 (1), 125–151.

Davidson, R.J., Tomarken, A.J., 1989. Laterality and emotion: an electrophysiological approach. Handb. Neuropsychol. 3, 419–441.

Desmet, P., 2003. Measuring emotion: development and application of an instrument to measure emotional responses to products. In: Funology. Springer, pp. 111–123. Retrieved from: http://link.springer.com/content/pdf/10.1007/1-4020-2967-5_12.pdf.

de Wijk, R.A., Kooijman, V., Verhoeven, R.H., Holthuysen, N.T., de Graaf, C., 2012. Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. Food Qual. Prefer. 26 (2), 196–203.

Dimberg, U., 1990. Facial electromyography and emotional reactions. Psychophysiology. Retrieved from: http://psycnet.apa.org/psycinfo/1991-11919-001.

Ferrarini, R., Carbognin, C., Casarotti, E.M., Nicolis, E., Nencini, A., Meneghini, A.M., 2010. The emotional response to wine consumption. Food Qual. Prefer. 21 (7), 720–725.

Gellhorn, E., 1970. The emotions and the ergotropic and trophotropic systems. Psychol. Forsch. 34 (1), 48–66.

Guo, F., Ding, Y., Wang, T., Liu, W., Jin, H., 2016. Applying event related potentials to evaluate user preferences toward smartphone form design. Int. J. Ind. Ergon. 54 (Suppl. C), 57–64. https://doi.org/10.1016/j.ergon.2016.04.006.

Helander, M.G., 2001. Theories and methods in affective human factors design. In: Smith, M.J., Salvendy, G., Harris, D., Koubek, R.J. (Eds.), Usability Evaluation and Interface Design, Vol. 1 of the Proceedings of HCI International 2001. Lawrence Erlbaum Associates, Mahwah NJ, pp. 357–361.

Hill, A.P., Bohil, C.J., 2016. Applications of optical neuroimaging in usability research. Ergon. Des 24 (2), 4–9.

Hodes, R.L., Cook, E.W., Lang, P.J., 1985. Individual differences in autonomic response: conditioned association or conditioned fear? Psychophysiology 22 (5), 545–560.

Hsiao, K.-A., Chen, L.-L., 2006. Fundamental dimensions of affective responses to product shapes. Int. J. Ind. Ergon. 36 (6), 553–564. https://doi.org/10.1016/j.ergon.2005.11.009.

Hu, S., Player, K.A., McChesney, K.A., Dalistan, M.D., Tyner, C.A., Scozzafava, J.E., 1999. Facial EMG as an indicator of palatability in humans. Physiol. Behav. 68 (1), 31–35.

Jenke, R., Peer, A., Buss, M., 2014. Feature extraction and selection for emotion recognition from EEG. Affective computing. IEEE Transactions on 5 (3), 327–339.

Khalid, H.M., 2004. Guest editorial: conceptualizing affective human factors design. Theor. Issues Ergon. Sci. 5 (1), 1–3.

King, S.C., Meiselman, H.L., 2010. Development of a method to measure consumer emotions associated with foods. Food Qual. Prefer. 21 (2), 168–177. https://doi.org/10.1016/j.foodqual.2009.02.005.

Kling, C., 1933. The role of the parasympathetics in emotions. Psychol. Rev. 40 (4), 368.

Kramer, A.F., 2006. Cognitive psychophysiology in ergonomics and human Factors. In: Karwowski, W. (Ed.), International Encyclopedia of Ergonomics and Human Factors, second ed. CRC Press, pp. 615–618.

Kreibig, S.D., 2010. Autonomic nervous system activity in emotion: a review. Biol. Psychol. 84 (3), 394–421.

Lane, R.D., Ahern, G.L., Schwartz, G.E., Kaszniak, A.W., 1997. Is alexithymia the emotional equivalent of blindsight? Biol. Psychiatr. 42 (9), 834–844.

Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O., 1993. Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology 30, 261.

Laparra-Hernández, J., Belda-Lois, J.M., Medina, E., Campos, N., Poveda, R., 2009. EMG and GSR signals for evaluating user's perception of different types of ceramic flooring. Int. J. Ind. Ergon. 39 (2), 326–332. https://doi.org/10.1016/j.ergon.2008.02.011.

Lee, M., Cho, G., 2009. Measurement of human sensation for developing sensible textiles. Human Fact. Ergon. Manufact. Service Industries 19 (2), 168–176. https://doi.org/10.1002/hfm.20144.

Liu, Y., Sourina, O., 2012. EEG-based valence level recognition for real-time applications. In: Cyberworlds (CW), 2012 International Conference on (Pp. 53–60). IEEE.

Liu, Y., Sourina, O., Nguyen, M.K., 2011. Real-time EEG-based emotion recognition and its applications. In: Transactions on Computational Science XII (Pp. 256–277). Springer. Retrieved from: http://link.springer.com/chapter/10.1007/978-3-642-22336-5_13.

Mauss, I.B., Robinson, M.D., 2009. Measures of emotion: a review. Cognit. Emot. 23 (2), 209–237. https://doi.org/10.1080/02699930802204677.

Mehrabian, A., 1980. Basic dimensions for a general psychological theory: implications for personality, social, environmental, and developmental studies. Retrieved from: http://philpapers.org/rec/MEHBDF.

Mehrabian, A., 1995. Framework for a comprehensive description and measurement of emotional states. Genet. Soc. Gen. Psychol. Monogr. Retrieved from: http://psycnet.apa.org/psycinfo/1996-93421-001.

Motte, D., 2009. Using brain imaging to measure emotional response to product appearance. In: Proceedings of International Conference on Designing Pleasurable Products and Interfaces, Compiegne, France.

Murphy, K.R., Jako, R.A., Anhalt, R.L., 1993. Nature and consequences of halo error: a critical analysis. J. Appl. Psychol. 78 (2), 218.

Nielsen, J., Levy, J., 1994. Measuring usability: preference vs. performance. Commun. ACM 37 (4), 66–75.

Niemic, C.P., Warren, K., 2002. Studies of emotion. A theoretical and empirical review of psychophysiological studies of emotion(Department of Clinical and

Social Psychology). JUR Rochester 1 (1), 15–19.
Park, C., Looney, D., Mandic, D.P., 2011. Estimating human response to taste using EEG. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (Pp. 6331–6334). IEEE. Retrieved from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6091563.
Paulhus, D.L., John, O.P., 1998. Egoistic and moralistic biases in self-perception: the interplay of self-deceptive styles with basic traits and motives. J. Pers. 66 (6), 1025–1060.
Paulhus, D.L., Vazire, S., 2007. The Self-report Method. In: Robins, R.W., Fraley, R.C., Krueger, R.F. (Eds.), Handbook of Research Methods in Personality Psychology. Guilford Press, New York, pp. 224–239.
Peck, E.M., Afergan, D., Jacob, R.J.K., 2013. Investigation of fNIRS brain sensing as input to information filtering systems. In: Proceedings of the 4th Augmented Human International Conference. ACM, New York, NY, USA, pp. 142–149. https://doi.org/10.1145/2459236.2459261.
Petrantonakis, P.C., Hadjileontiadis, L.J., 2010. Emotion recognition from EEG using higher order crossings. Inf. Technol. Biomed. IEEE Transact. 14 (2), 186–197.
Porcherot, C., Delplanque, S., Raviot-Derrien, S., Le Calvé, B., Chrea, C., Gaudreau, N., Cayeux, I., 2010. How do you feel when you smell this? Optimization of a verbal measurement of odor-elicited emotions. Food Qual. Prefer. 21 (8), 938–947.
Provins, K.A., Cunliffe, P., 1972. The relationship between EEG activity and handedness. Cortex 8 (2), 136–146.
Qie, N., Rau, P.-L.P., Deng, J., 2017. Emotions evoked by traditional Chinese herbs for cosmeceuticals. Int. J. Affect. Engin. 16 (2), 57–62. https://doi.org/10.5057/ijae.IJAE-D-16-00018.
Quintana, D.S., Guastella, A.J., Outhred, T., Hickie, I.B., Kemp, A.H., 2012. Heart rate variability is associated with emotion recognition: direct evidence for a relationship between the autonomic nervous system and social cognition. Int. J. Psychophysiol. 86 (2), 168–172. https://doi.org/10.1016/j.ijpsycho.2012.08.012. Retrieved from.
Ramirez, R., Vamvakousis, Z., 2012. Detecting emotion from EEG signals using the emotive epoc device. In: Brain Informatics. Springer, pp. 175–184. Retrieved from: http://link.springer.com/chapter/10.1007/978-3-642-35139-6_17.
Russell, J.A., 1979. Affective space is bipolar. J. Pers. Soc. Psychol. 37 (3), 345.
Scherer, K.R., 2005. What are emotions? And how can they be measured? Soc. Sci. Inf. 44 (4), 695–729.
Schmidt, L.A., Trainor, L.J., 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. Cognit. Emot. 15 (4), 487–500.
Sourina, O., Liu, Y., 2011. A fractal-based algorithm of emotion recognition from EEG using arousal-valence model. In: Biosignals, pp. 209–214. Retrieved from: http://ntu.edu.sg/home/eosourina/Papers/OSBIOSIGNALS_66_CR.pdf.
Thorndike, E.L., 1920. A constant error in psychological ratings. J. Appl. Psychol. 4 (1), 25–29.
Trimmel, M., Fairclough, S., Henning, R., 2009. Psychophysiology in ergonomics. Appl. Ergon. 40 (6), 963.
van den Broek, E.L., Westerink, J.H.D.M., 2009. Considerations for emotion-aware consumer products. Appl. Ergon. 40 (6), 1055–1064. https://doi.org/10.1016/j.apergo.2009.04.012.
Valenza, G., Citi, L., Lanatá, A., Scilingo, E.P., Barbieri, R., 2014. Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. Sci. Rep. 4. Retrieved from http://www.nature.com/srep/2014/140518/srep04998/full/srep04998.html.
Warrenburg, S., 2005. Effects of fragrance on emotions: moods and physiology. Chem. Senses 30 (Suppl. 1), i248–i249.
Zentner, M., Grandjean, D., Scherer, K.R., 2008. Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8 (4), 494.