That sounds sweet: using cross-modal correspondences to communicate gustatory attributes

Klemens M. Knöferle
BI Norwegian Business School
Andy Woods
Florian Käppler
Charles Spence

This is the accepted, refereed and final manuscript to the article published in

Psychology and Marketing, 32(2015)1:107-120

Publisher's version available at http://dx.doi.org/10.1002/mar.20766

That Sounds Sweet: Using Crossmodal Correspondences to Communicate Gustatory Attributes

Klemens M. Knoeferle

BI Norwegian Business School

Andy Woods

Xperiment

Florian Käppler

Trossingen University of Music

Charles Spence

University of Oxford

Author Notes

Klemens M. Knoeferle, Department of Marketing, BI Norwegian Business School, Norway; Andy Woods, Xperiment, Switzerland; Florian Käppler, Department of Music Design, Trossingen University of Music, Germany; Charles Spence, Department of Experimental Psychology, University of Oxford, United Kingdom.

Correspondence concerning this article should be addressed to Klemens Knoeferle, Department of Marketing, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway, +47 464 10 932, klemens.knoferle@bi.no.

Acknowledgments: Parts of this research were funded by the Swiss National Science Foundation during the first author's post-doctoral studies at the University of Oxford (grant PBSGP1_141347).

Abstract

Building on existing research into sound symbolism and crossmodal correspondences, this article proposes that crossmodal correspondences—systematic mappings between different sensory modalities—can be used to communicate non-musical, low-level sensory properties such as basic tastes through music. A series of three experiments demonstrates that crossmodal correspondences enable people to systematically encode basic taste properties into parameters in musical space (Experiment 1), and that they are able to correctly decode basic taste information embedded in complex musical compositions (Experiments 2 and 3). The results also suggest some culture-specificity to these mappings, given that decoding performance, while still above chance levels, was lower in Indian participants than in those from the United States (Experiment 3). Implications and potential applications of these findings are discussed.

Keywords: crossmodal correspondences: multisensory processing; music, psychoacoustics; sound, taste, cross-cultural

That Sounds Sweet: Using Crossmodal Correspondences to Communicate Gustatory Attributes

Over recent years, marketers have become increasingly interested in the study of sound symbolism; that is, the question of the extent to which the acoustic parameters of various kinds of sounds can communicate meaning (Argo, Popa, & Smith, 2010; Coulter & Coulter, 2010; Doyle & Bottomley, 2011; Klink, 2000; Klink & Athaide, 2012; Kuehnl & Mantau, 2013; Lowrey & Shrum, 2007; Luna, Carnevale, & Lerman, 2013; Shrum, Lowrey, Luna, Lerman, & Liu, 2012; Yorkston & Menon, 2004). For instance, when participants had to select one of two artificial product names (differing only in the vowel sound) for a new product, they preferred product names whose vowel sounds connoted properties that were congruent with the product's attributes (e.g., for a large car, they selected names with large-sounding, rather than small-sounding vowels; Lowrey & Shrum, 2007).

Surprisingly, research into the topic of sound symbolism in marketing has focused exclusively on phonetic symbolism (semantic meaning conveyed by means of speech sounds). In contrast, the question of the extent to which (and under what conditions) meaning is conveyed by music has received only scant

attention. The present research was designed to address this gap in the literature by investigating whether music can communicate basic taste properties; that is, whether individuals reliably associate sounds varying in their psychoacoustic and musical parameters with gustatory attributes. Specifically, crossmodal correspondences between the parameters of music and the verbal representations of basic tastes (e.g., sweet, sour, salty, bitter) were tested, and the cross-cultural robustness of the effects assessed.[1]

To motivate both the research question and its relevance to the marketing community, the authors draw on earlier research into sound symbolism and crossmodal correspondences. In terms of the former, existing research on sound symbolism suggests that music can communicate specific, extramusical meaning. In terms of the latter, past research on crossmodal correspondences indicates that specific auditory and gustatory dimensions are reliably associated with each other. Next, a series of three experiments is reported, which is designed to test whether individuals reliably match musical and gustatory dimensions. Experiment 1 replicates and extends the auditory-gustatory mappings that have been reported recently. Specifically, Experiment 1 tests whether musical laypersons encode different basic taste labels into systematically different regions of musical space, operationalized through a range of psychoacoustic parameters. Experiment 2 tests whether newly created soundtracks are reliably assigned to given taste categories by musical laypersons. Experiment 3 examines whether cultural background affects sound-taste matching ability; that is, whether there are significant differences in the extent to which crossmodal correspondences emerge across cultures with different musical backgrounds. Finally, the implications of these findings, potential practical applications, and some intriguing avenues for future research are discussed.

Background

Sound Symbolism and Musical Meaning

The notion of music as a carrier of semantic meaning has received empirical support from a number of studies measuring behavioural outcomes (Watt & Quinn, 2007; Zhu & Meyers-Levy, 2005) and measures of brain activity (Daltrozzo & Schön, 2009a, 2009b; Koelsch, et al., 2004; Painter & Koelsch, 2011). In a forced-choice task, Watt and Quinn (2007) demonstrated that listeners will robustly map musical qualities onto a number of non-musical qualities, with performance being best for human qualities (e.g., male-female, evil-good), possibly due to shared patterns of temporal and dynamic change that are present in both music and individual human behaviour.

The question remains as to whether and how the conceptual units in music can refer to low-level sensory concepts, such as concepts of basic tastes. According to one recent account, sound symbolism may be just a special case of the more general phenomenon of crossmodal correspondences (see Spence, 2011; Walker & Walker, 2012), which will be discussed in more detail in the next section.

Crossmodal Correspondences Between Sound and Taste

The term “crossmodal correspondences” (also “crossmodal analogies”, “synesthetic correspondences”) refers to the phenomenon that even non-synesthetic individuals often map basic stimulus properties from different sensory modalities onto each other in a manner that is surprisingly consistent (for a review, see Spence, 2011). For instance, angular shapes are reliably matched to higher (rather than lower) auditory pitch (Marks, 1987; Walker, 2012). Congruence (as opposed to incongruence) of such dimensions has been found to speed up response latencies across several experimental paradigms including speeded detection, classification, and visual search (Evans & Treisman 2010; Gallace & Spence 2006; Marks 1987; Walker 2012), indicating facilitation / interference processes operating on an automatic, perceptual level (Parise & Spence 2013). Furthermore, audiovisual correspondence effects have been demonstrated in primates (Ludwig, Adachi, & Matsuzawa 2011), suggesting that linguistic associations cannot be the only explanation of the phenomenon.

Specific to audition and taste, recent cognitive psychology and neuroscience research has demonstrated that people will reliably match specific (psycho-)acoustic and musical parameters with different tastes, flavors, and oral-somatosensory food-related experiences (Crisinel, et al., 2012; Crisinel & Spence, 2010a, 2010b, 2011, 2012a, 2012b; Mesz, Sigman, & Trevisan, 2012; Mesz, Trevisan, & Sigman, 2011; Simner, Cuskley, & Kirby, 2010). As these recent findings have been comprehensively discussed elsewhere (Knöferle & Spence, 2012), the following review will focus only on the most important aspects.

For example, Crisinel and Spence tested the strength of the crossmodal associations that exist between the pitch of musical sounds and sour versus bitter tastes (Crisinel & Spence, 2009), and sweet versus salty tastes (Crisinel & Spence, 2010b) using the Implicit Association Test. The results suggested that sweet and sour tastes are associated with high pitch, whereas bitter tastes are associated with low pitch. In a related study, Crisinel and Spence (2010a) highlighted the existence of a number of crossmodal correspondences between the sounds of various musical instruments (i.e., sounds that differ in timbre) and basic tastes. For instance, these researchers found that bitter and sour tastes were reliably mapped to the sound of the trombone (evaluated as rather unpleasant by participants), while sweet tastes are typically mapped to piano sounds (evaluated as rather pleasant by participants).

In another study, free piano improvisations from professional musicians on the theme of basic taste words (e.g., “salty”) reliably resulted in distinct musical properties. For example, the term "salty" was musically expressed using low pitch and staccato (i.e., short, abrupt) phrasings, while improvisations on the word "bitter" were expressed using low pitch and legato (i.e., smooth, flowing) phrasings (Mesz, et al., 2011). In a subsequent experiment, musically naïve listeners were able to reliably map the pianists' musical improvisations to the four basic tastes. A follow-up study subsequently demonstrated that taste-congruent music produced by means of a computer algorithm is reliably matched to basic tastes above chance level by musical laypersons (Mesz, et al., 2012).

Are Crossmodal Correspondences Between Sound and Taste Culture-Specific?

The question of what mechanisms drive the ability to match auditory and gustatory dimensions has not been empirically answered, although a range of potential drivers has been discussed in earlier research (e.g., Knöferle & Spence, 2012; Spence, 2011, 2012). In particular, the extent to which cultural aspects shape crossmodal correspondences between sounds and tastes is unclear. Would participants from different cultural and hence musical backgrounds share similar sound-taste associations? On the one hand, sound-taste correspondences may be primarily driven by culture-specific linguistic or musical metaphors (e.g., using the adjective "sweet" to characterize a bundle of expressive properties in Western music; see also Eitan & Timmers, 2010). On the other hand, however, crossmodal correspondences between sound and taste may be based on more fundamental mechanisms (e.g., learning processes that are unaffected by culture, or the innate neural architecture; for a discussion see Spence & Deroy, 2012). Indeed, certain other types of crossmodal correspondences (e.g. the correspondence between redness and sweet taste) appear to result from the internalization of the statistical properties of the environment (i.e., red fruit are often ripe and sweet; Spence, Levitan, Shankar, & Zampini, 2010) in a Hebbian type of learning (Hebb, 1949)—a mechanism that comes with apparent benefits for object recognition and most likely individual survival. The adaptive reasons for robust sound-taste associations, in contrast, are less clear (similar to sound-smell correspondences; see Deroy, Crisinel, & Spence, 2013). Alternatively, auditory-gustatory associations have been proposed to stem from embodied learning processes. For example, the mapping of sweet taste and high pitch may be based on the innate orofacial approach/avoidance gestures and associated speech sounds with which human infants respond to sweet versus bitter tastes (Knöferle & Spence, 2012; Scherer, 1986; Spence, 2012). Typical reactions to bitter tastes include protruding the tongue outwards and downwards, resulting in lower-frequency speech sounds when exhaling, while typical reactions to sweet tastes involve outwards and upwards tongue positions, which result in higher-frequency sounds.

Consequently, if crossmodal correspondences between sounds and tastes were to be observed in Western, but not in non-Western samples, this would indicate that associations formed during an individual's musical socialization constitute the main drivers of the phenomenon, and therefore provide strong evidence for a cultural account of sound-taste correspondences. Alternatively, however, if such crossmodal correspondences were to emerge irrespective of the participants' musical background, this may point to a more fundamental process as a driver of sound-taste mappings.

Importantly, the question of whether sound-taste correspondences are influenced by cultural background is very relevant to marketing. For marketing professionals who develop media materials involving music, it is crucial to know the extent to which crossmodal associations and sound symbolism effects are shared among different populations and markets (cf. intercultural studies on phonetic symbolism, e.g., Kuehnl & Mantau, 2013; Shrum, et al., 2012). Depending on the answer to this question, it may be more appropriate for marketers to adopt either a market-specific or a global music strategy.

Experiment 1: What Taste Does That Sound Have?

Experiment 1 examines whether and to what extent individuals encode different basic taste labels into different regions of musical space. In contrast to Mesz and colleagues' (2011) first study, Experiment 1: a) focuses on the question of whether layperson participants, rather than professional musicians, reliably encode basic tastes into sounds; and b) includes important psychoacoustic attributes (e.g., psychoacoustic sharpness and roughness) that are amongst the most salient parameters of auditory perception (Fastl & Zwicker, 2007).

Method

An online within-participants experiment was conducted, in which participants from the UK (*N* = 39; 21 female; mean age = 31.7 years) were asked to match selected auditory parameters of an experimental sound to basic taste words. Before starting the main study, the participants were required to enter the correct response to an acoustically presented question to ensure sound playback was active, and to set a comparable (i.e., comfortable) sound level. Then, on each trial, the participants were presented with one of four basic taste words (sweet, sour, salty, bitter) and a seven-point slider scale. By moving the slider handle, the participants were able to adjust a short chord progression (three G7 chords leading to a C major chord anchored at C_5 = 523.25 Hz) in terms of one of six auditory properties. More specifically, they could use the slider to increase or decrease (in seven steps) the sound's attack, discontinuity, pitch, roughness, sharpness, or speed (for a short explanation of these concepts see Knöferle & Spence, 2012).[3] A sound that corresponded to the selected scale value was played back every time the participant clicked on the slider handle, one of the scale values, or moved the scale handle to a new value. The participants were instructed to adjust the slider so that the sound best matched the basic taste. The instructions of the experiment made clear that the presented words would all refer to basic tastes.

The participants confirmed their final selection by clicking on a button that initialized the next trial. Each participant was presented with all 24 possible combinations of sounds and taste words (6 auditory properties × 4 taste words) in random order. In this study, special attention was given to equalize the loudness of all of the auditory stimuli in order to rule out a confounding of auditory roughness and pitch with auditory loudness (e.g., without loudness equalization, changing a sound's pitch inevitably changes its loudness; for a discussion see Knöferle & Spence, 2012). This was accomplished by computing averaged loudness values for each of the auditory stimuli as described in ISO 532 B, and then adjusting all sounds to the smallest obtained loudness value.

Results

Systematic patterns emerged in participants' sound-taste word mappings (see Figure 1). All subsequent analyses use listwise deletion for missing data; all pairwise comparisons use the Bonferroni correction. A repeated-measures ANOVA revealed a statistically significant effect of taste word on pitch ratings, $F(3, 96) = 8.75$, $p < .001$.

Post hoc analyses demonstrated that compared to bitter taste ($M = 3.42$, $SD = 2.14$), sour taste ($M = 5.24$, $SD = 1.89$, $p = .002$), and sweet taste ($M = 5.39$, $SD = 1.56$, $p = .002$), were mapped onto a significantly higher pitch. The salty taste ($M = 4.52$, $SD = 1.75$) ranged between bitter and sour tastes in terms of its pitch height, being marginally significantly higher in pitch than the bitter taste ($p = .08$). Second, the results indicate that people consistently map auditory roughness (Terhardt, 1974) onto basic tastes, $F(3, 96) = 25.68$, $p < .001$. Post-hoc tests revealed that participants selected the lowest roughness values for sweet taste words ($M = 1.97$, $SD = 1.85$), significantly higher values for salty taste ($M = 4.00$, $SD = 2.03$, $p = .002$), and, compared to salty, significantly higher values for sour ($M = 5.09$, $SD = 1.79$, $p = .024$) and bitter tastes ($M =$

5.27, *SD* = 1.84, *p* = .018). The effect of discontinuity was significant, $F(3, 90) = 14.74$, $p < .001$; the sweet taste was mapped onto sounds that were lower in discontinuity (*M* = 2.90), whereas sour (*M* = 5.28), salty (*M* = 5.34), and bitter (*M* = 5.28) tastes were all mapped onto sounds having a higher discontinuity (all *p*s < .001). A significant difference also emerged for musical tempo, $F(3, 99) = 4.293$, $p = .007$, with sour tastes (*M* = 5.39) resulting in significantly higher average tempo ratings than the bitter taste (*M* = 4.03, *p* = .02, all other pairwise comparisons non-significant). Finally, the overall effect of auditory sharpness was significant, $F(3, 99) = 3.03$, $p = .033$, with sour tastes being mapped onto sharper sounds as compared to the other basic tastes. However, pairwise comparisons revealed that this difference was, at best, marginally significant between the sour taste (*M* = 5.44) and the bitter taste (*M* = 4.12, *p* = .063). Attack/onset time of the sounds was not reliably linked to any of the basic tastes ($F < 1$). The observed response patterns clearly indicate that participants' response behaviour was systematic for certain, but not all of the auditory dimensions.

Insert Figure 1 here.

Experiment 2: What Taste Is That Sound?

While Experiment 1 focused on encoding taste labels into musical space, Experiment 2 was designed to test whether participants would be able to decode musical meaning that refers to basic tastes and that is systematically embedded in the low-level properties of musical compositions.

Method

Following the methodology popularized in the early literature on phonetic symbolism (Köhler, 1929, 1947; Sapir, 1929), participants were asked to match a number of soundtracks to basic taste labels. Compared to related previous studies, Experiment 2 introduced some critical methodological improvements. Whereas previous studies utilized musical compositions that varied rather freely within musical space (e.g., by using different MIDI instruments across compositions, Mesz, et al., 2012; by using completely different musical material, Mesz, et al., 2011), Experiment 2 used music that exclusively varied along a series of (psycho-)acoustic dimensions that were systematically manipulated. This resulted in a more rigorous test of whether participants are able to match music to basic tastes, since differences in participants' mapping performance can be solely ascribed to differences in the manipulated low-level music properties. The chosen design has the added benefit that low-level features of music can be varied in virtually all types of music. They are largely independent of musical genre (although see the General Discussion for potential exceptions), and thus can be implemented across different marketing contexts, increasing the relevance of the current research for marketing practice.

A sound branding agency was instructed to create four pieces of music by systematically varying the low-level properties of a 30-second piece of music consisting of several background (synthesizer bass and pads) and foreground elements (piano and arpeggiated synthesizer). Based on auditory-gustatory correspondences identified in Experiment 1 and previous studies, auditory pitch, sharpness, roughness, consonance, and discontinuity were manipulated (see Table 1 for detailed information concerning the manipulations). Loudness was held constant across the four compositions using the same method as in Experiment 1.

Insert Table 1 here.

In an online experiment, 61 participants from the US recruited via Amazon Mechanical Turk (14 female, mean age = 31 years) assigned the four musical compositions to four basic tastes words (bitter, sweet, sour, and salty) in a forced-choice task. The four musical compositions were represented by four simultaneously-presented interactive on-screen sound objects (randomized positioning) that could be played back by clicking on them. The participants were instructed to play back all of the sounds as often as they liked and then to drag-and-drop each of the sound objects onto one of four labelled taste word bins. Each bin could only be assigned to one sound object, and all sound objects had to be assigned to a bin. The participants could change their selections without time constraints and until they were satisfied with their choices.

Results

The participants decoded the correct taste word for each piece of music at a level that was well significantly above chance. 17 out of 61 participants showed perfect performance and correctly assigned all four sounds. The probability for correct mappings under the assumption of purely random response behaviour (chance) is 9/24 or $p = .375$ for zero correct responses, 8/24 or $p = .333$ for one correct response, 6/24 or $p = .25$ for two correct responses, and 1/24 or $p = .042$ for four correct responses. The probability for three correct responses is $p = 0$, since three correct responses leaves only one possible outcome for the fourth mapping, which then inevitably will be correct. As Figure 2 illustrates, the observed probabilities are thus systematically lower than expected under the assumption of random choice behaviour for zero and one correct matches and systematically higher for four correct matches.

Insert Figure 2 here.

On average, the participants assigned 1.90 (*SD* = .19) of the four sounds to the correct taste word. This is nearly twice the number that would be expected under the assumption of random mappings, which is 1. Participants' performance was positively correlated with the amount of time they used to complete the experimental task ($R = .285$, $p = .026$), which, together with the fact that an online sample was used, suggests that the performance observed under more controlled laboratory conditions might have been higher. Performance was not significantly correlated with the amount of active musical experience, with the sex, or with the age of the participants (all *p*s > .39).

To test whether the observed response distribution and its parameters differed significantly from a purely random response distribution, a Monte Carlo simulation was performed (as the response distribution was non-Gaussian; see Mesz, 2012). Specifically, 10^6 experiments were simulated, each containing $N = 61$ random measures of matching performance, thus modelling random mapping behaviour. None of the 10^6 Monte Carlo samples resulted in an average performance measure that was larger than 1.9, which is the average performance in the observed data (the maximum simulated value was 1.62, see Figure 3). This suggests that the probability of the observed results being the consequence of purely random response behaviour is $p < 10^{-6}$.

Insert Figure 3 here.

The observed matching performance was most accurate for the sweet taste, slightly less accurate for the salty taste, and least accurate for the bitter and sour tastes (see Table 2). This specific pattern of performance may at least partially be explained by the finding that many English speakers regularly confuse bitter and sour tastes (O'Mahony, Goldenberg, Stedmon, &

Alford, 1979). Altogether, Experiment 2 provides robust evidence suggesting that individuals are able to decode music in terms of gustatory concepts above chance level.[4]

Insert Table 2 here.

Experiment 3: Does Sound-Taste Matching Generalize Across Cultures?

The goal of Experiment 3 was to test whether the strength of the sound-taste associations found in Experiment 2 would vary as a function of cultural and musical background. As described above, the ability to decode musical meaning in terms of basic tastes may be based on associative learning processes and culture-specific metaphors (e.g., “soft and flowing sounds are sweet”), dependent on the exposure to a specific (musical) language and tradition. Alternatively, however, the ability could be a consequence of non-cultural factors, such as the common neural coding of a particular stimulus dimension (e.g., intensity across sensory modalities is represented by increased neural firing; Spence, 2011) or embodied experiences that occur across cultures. Therefore, if the effect was culture-dependent, Western participants (for whom the effect has been shown in Experiment 2) would be expected to perform better in the matching task than participants from non-Western cultures, whose musical associations are shaped by a different musical environment (Morrison & Demorest, 2009; Murray & Murray, 1996; Stevens, 2012). If, on the other hand, the effect was driven by extra-cultural factors, decoding performance should be equivalent across Western and non-Western cultures.

Method

309 participants from the US (122 female, mean age = 32 years) and 207 participants from India (72 female, mean age = 33 years) were recruited through Amazon Mechanical Turk and

took part in this study in exchange for a payment of 0.20 US dollars. An Indian sample was selected because the musical socialization of Indian participants differs considerably from Western participants. For example, while traditional Indian music is based on a microtonal system, traditional Western music uses a tonal system with 12 semitone intervals (Agarwal, Karnick, & Raj, 2013). As a consequence of such profound structural differences, the two musical cultures likely entail different musical metaphors. The procedure and stimuli were identical to the one used in Experiment 2 (forced-choice matching of the four soundtracks to the four basic taste labels). Two of the participants from the US and four from India were excluded as they had completed the study without listening to all the musical stimuli.

Results

61 out of 307 or 20% of all US participants matched all sounds to the “correct” tastes. In contrast, only 17 out of 203 or 8.4% of all Indian participants matched all sounds to the “correct” tastes. Both the US and the Indian sample exceeded chance level in terms of the number of correct mappings as computed in Experiment 2 ($M_{\text{US}} = 1.75$, $SD = 1.32$; $M_{\text{India}} = 1.38$, $SD = 1.10$). Bootstrapping analysis as described in Experiment 2 revealed that both sample means were significantly higher than expected under chance matching ($M_{\text{chance}} = 1.00$, 99% CI [0.69, 1.34], $p < .01$).

To test whether the likelihood of correct sound-taste mappings differed as a function of country and sound stimulus, a generalized linear mixed-effects model (GLMM) with binomial error distribution and a log link-function was used. Such a model accommodates both our binary outcome variable (correct/incorrect matching of any given sound) and correlated error terms that likely arise from the repeated measurement structure of our data (Fitzmaurice, Laird, & Ware, 2004). All statistics were performed in the R statistical programming environment using the glmer function of the lme4 package (Bates & Sarkar, 2013).

The data consisted of four sound-taste mappings from each of the 510 participants, resulting in a total of 2040 observations. The original dependent variable (taste label: bitter, salty, sour, sweet; as a response to a given soundtrack) was transformed into a binary variable (correct vs. incorrect match). Three models were estimated, differing with regard to the included predictors. The first model contained dummy variables for the four sounds and a control variable measuring the amount of musical experience in order to control for individual differences in musical training that might otherwise drive participants' matching performance. A random intercept was included to account for the likely correlated error terms of observations from individual subjects (i.e., repeated measures). The first model yielded significant coefficients for all included variables and a significant increase in fit compared to a "null model" (intercept-only model) without any fixed effects ($\chi^2 = 106.81$, $df = 4$, $p < .001$).

The second model in addition contained a dummy variable for country US (but no interaction term). This model yielded significant coefficients for the country US variable, for the salty sound (marginal), for the sour sound, for the sweet sound, and music experience. The second model also yielded a significant increase in fit compared to the first model as indicated by fit statistics and a likelihood-ratio test ($\chi^2 = 12.06$, $df = 4$, $p = .02$).

The third model contained the terms of the previous model plus interaction terms of the country and sound dummy variables in order to test whether specific sounds might have been matched more or less accurately across the two countries. This model yielded significant coefficients for the sour sound, the sweet sound, music experience, and importantly, the interaction between country US and the sweet sound (marginal). None of the other variables reached significance ($p > .1$). While the improvement in model fit of this third model compared to the second model did not reach significance ($\chi^2 = 4.17$, $df = 3$, $p = .24$), the significant interaction between country and sound is theoretically informative, indicating that depending on the country, some sounds were matched more easily than others. The summarized results and properties associated with each of the three models are included in Table 3.

Insert Table 3 here.

The direction of the country coefficient in model 2 suggests that the participants from the US, controlling for music experience, were significantly better at matching the sounds to the correct taste words compared to Indian participants. Further, the direction of the sound coefficients in models 2 and 3 suggest that independent of country, the four sounds led to different matching performance: Compared to bitter sounds, sweet sounds were significantly easier to match, while sour and salty sound were significantly harder to match. Adding the country × sound interaction terms in model 3, while not resulting in a significantly better model fit compared to model 2, revealed a significant interaction between country US and the sweet sound. This indicates that the significant country effect in model 2 might be mainly driven by a higher likelihood of US participants to correctly decode the sweet sound as “sweet”.

To further explore the role of musical experience, an additional model was run to test whether musical expertise moderated the effect of country; that is, whether only those Indian participants with high musical expertise might match the performance of US participants. To this end, an additional interaction term between country and musical expertise was included in model 2; however, the associated parameter was not significant ($p > .38$).

Interestingly, Indian participants took longer to complete the study than participants from the US ($M_{\text{India}} = 281$, $M_{\text{US}} = 152$ s, $t(508) = -9.70$, $p < .001$). One possible explanation for this difference in completion time is that Indian participants likely would have found the sound-taste matching task more difficult than American participants (in line with a cultural account of crossmodal correspondences), requiring them to pay more attention and thus take longer to complete the task.

General Discussion

Summary

The present results demonstrate that musical compositions that match certain basic taste qualities can be systematically created by using crossmodal correspondences (see Spence, 2011, for a review). Experiment 1 demonstrated that participants can systematically encode basic tastes into low-level (psycho-)acoustic attributes. Different basic taste words were mapped to different locations in musical space. For example, the bitter taste was mapped to rough, slow, and low-pitched sounds, whereas sweet tastes were mapped to high-pitched, smooth, and continuous sounds. Sour taste was characterized by higher sharpness and speed, salty taste by intermediate values on all auditory parameters. Experiment 2 tested whether people are able to systematically match certain compositions with their appropriate referents (e.g., the sweet soundtrack with the sweet taste etc.). Indeed, the participants' performance was significantly higher than chance level. In Experiment 3, matching performance varied as a function of cultural background. Even when controlling for musical expertise, participants from the US performed significantly better than the Indian participants, thus indicating that cultural-specific associations may at least partially be driving the crossmodal correspondences between sound and taste. However, on average, Indian participants still performed above chance level, which may be regarded as evidence for crossmodal associations driven by more fundamental mechanisms (e.g., culture-independent statistical learning, embodied, or even innate links). Consistent with earlier research (Mesz, et al., 2011), the results reported in the current research further suggest that crossmodal correspondences between sound and taste may be bi-directional—taste words were first translated into musical parameters (Experiment 1) and the resulting compositions then successfully decoded by participants (Experiments 2 and 3).

It is important to note that the forced-response format of Experiments 1-3 does not impair the internal validity of the studies. Specifically, since the participants were forced to select one of several sounds in response to a taste word (Experiment 1) or one of several taste words in response to a sound (Experiments 2 and 3), one might argue that their choices were driven by some kind of demand effect rather than their own perceptual or semantic associations. However, such an account fails to explain the emergence of robust and convergent sound-taste mappings across individuals; that is, it cannot explain why the participants would, for example, reliably match sour taste to higher pitch than bitter taste. Therefore, demand characteristics do not seem applicable here. Still, the forced-choice design does not allow us to assess whether the mappings occurred spontaneously and automatically or as a result of deliberate thought (indeed, it is worth noting that the participants in earliery studies described pitch-taste mapping as effortful and challenging; Holt-Hansen, 1968; Rudmin & Cappelli, 1983).

Theoretical Implications

The present research contributes to a better understanding of crossmodal correspondences, and of sound symbolism (which may be regarded as a subset of crossmodal correspondences) in several ways. First, the present study adds to the growing body of research on sound symbolism in marketing and consumer psychology. To the best of the authors' knowledge, the study of sound symbolism has been limited to phonetic symbolism in the marketing literature (Argo, et al., 2010; Coulter & Coulter, 2010; Doyle & Bottomley, 2011; Klink, 2000; Klink & Athaide, 2012;

Kuehnl & Mantau, 2013; Lowrey & Shrum, 2007; Shrum, et al., 2012; Yorkston & Menon, 2004). The current study extends this narrow focus to musical stimuli, suggesting that sound symbolism can affect a much wider range of contexts and applications than previously assumed. Second, the present findings contribute to the related literature on crossmodal correspondences by providing additional evidence of shared underlying properties between the auditory and gustatory modality (Knöferle & Spence, 2012; Spence, 2012). Finally, the present research informs the ongoing debate about the extent to which crossmodal correspondences are “strong” (innate and automatic) or “weak” (learned and culture-specific) by providing evidence for auditory-gustatory correspondences in a non-Western sample.[5] The present findings indicate that auditory-gustatory correspondences may emerge due to a combination of culture-dependent and culture-independent factors.

The pattern of results in Experiment 1 is also informative regarding specific potential mechanisms that may drive crossmodal correspondences. First, a semantic matching account would predict that auditory and gustatory stimuli are matched based on learned semantic associations, that is, extensive bidirectional cross-activation among dimensions of connotative meaning (Walker, 2012). For example, “sweet” is a descriptor applicable to tastes and, metaphorically in Western culture, also to sounds. Accordingly, auditory cues featuring acoustic parameters that are, metaphorically, considered as “sweet” (such as low roughness and discontinuity) may be matched with sweet tastes, as observed in the current study. However, similarly, sweet tastes may be matched with higher pitch due to the semantic proximity between a pleasant taste that results in a “high” mood (Meier & Robinson, 2004) and the spatial metaphor for pitch description (low pitch – high pitch; Spence, 2011).

Second, sounds and tastes may be mapped onto each other based on their hedonic value. This hedonic account would predict that individuals match tastes that are perceived to be unpleasant (e.g., bitter) with sounds that are less pleasant, and more pleasant tastes (e.g. sweet) with sounds which are liked more. However, Crisinel and Spence (2012) by testing those who liked or disliked dark chocolate showed that pleasantness matching cannot fully explain the mapping between sweet (bitter) taste and high (low) pitch (see also Belkin, Martin, Kemp, & Gilbert, 1997). In the current study, hedonic relationships appeared to play an important role for the observed sound-taste mappings. The relative mapping of sweet taste onto low values of auditory roughness, for example, can be explained by the fact that sweet taste (in moderate intensity) is a pleasant and rewarding stimulus (Moskowitz, Kluter, Westerling, & Jacobs, 1974) and auditory roughness is negatively correlated with sensory pleasantness (Fastl & Zwicker, 2007). However, taste mappings in another auditory dimension that is strongly correlated with pleasantness (i.e., auditory sharpness) did not follow a hedonic matching account. Here, participants mapped the more unpleasant bitter taste onto more pleasant sounds that were low in auditory sharpness, but the sour sound to more unpleasant sounds with higher auditory sharpness. Therefore, the mapping must be driven by another, as yet unknown, principle.

Third, crossmodal correspondences may be driven by statistical co-occurrences in the environment that are implicitly or explicitly learned by individuals. Put differently, many crossmodal correspondences may simply reflect statistical (from the perspective of the consumer: learned) correlations between sensory features found in nature or the marketplace (Fiebelkorn, Foxe, & Molholm, 2010, 2012; Spence, 2012). For example, the fact that people associate larger objects with lower-pitched sounds, or higher-pitched sounds with smaller objects can be explained by the fact that in nature larger resonance bodies (e.g., in objects or animals) generally generate lower pitched sounds than smaller resonance bodies. With regard to auditory-gustatory correspondences, pitch and taste have been shown (both previously and in the current study) to share an implicit association that cannot be explained through a hedonic matching account alone

(Crisinel & Spence, 2012b). According to a statistical co-occurrence account, one speculative explanation for this finding seems to be a statistical learning account based on the innate orofacial approach/avoidance gestures and associated speech sounds with which human infants respond to sweet versus bitter tastes (Knöferle & Spence, 2012; Scherer, 1986; Spence, 2012). Typical reactions to bitter tastes include protruding the tongue outwards and downwards, resulting in lower-frequency speech sounds when exhaling, while typical reactions to sweet tastes involve outwards and upwards tongue positions, which result in higher-frequency sounds.

Importantly, some of the present results cannot readily be explained by any of the mechanisms proposed above. For example, why would sour tastes be consistently mapped to a higher musical tempo? This particular mapping is neither accounted for by a hedonic, statistical, linguistic/semantic, or intensity correspondence. One putative explanation would be that more aversive tastes might generally trigger faster responses (e.g., potentially survival-critical avoidance behaviours; Glendinning, 1994) and therefore induce a higher level of activity than pleasant tastes (as shown in the domain of olfaction; Boesveldt, Frasnelli, Gordon, & Lundström, 2010). However, not all of the more aversive basic tastes (i.e., only sour, but not bitter taste) were linked to high musical tempo in the present study. Future research is certainly needed to further disentangle the factors influencing sound-taste correspondences.

The present findings also contribute to the literature on cross-cultural differences in the perception of advertising. Previous research has, for example, investigated the role of visual factors (An, 2007; Zhou, Zhou, & Xue, 2005) and music (Murray & Murray, 1996; Tavassoli & Lee, 2003) in advertising across different cultures. While prior studies focused on musical surface features (e.g., genre or the meaning of lyrics; Murray & Murray, 1996; presence vs. absence of music; Tavassoli & Lee, 2003), the present study extends this line of research by investigating how meaning that is encoded in the very musical structure is differentially perceived across cultures. Our findings may also have implications for the processing of advertising messages: Past research has found non-verbal auditory cues to interfere more with message processing when paired with English rather than Chinese advertisement messages—likely due to the fact that alphabetical languages rely more on auditory processing and the subvocalizing of words, while logographic languages depends more on visual processing (Tavassoli & Lee, 2003). Given that music has been shown to carry semantic meaning (Koelsch, et al., 2004), a possibility is that this interference effect of music on alphabetical language processing will be reduced if the meaning of the music is congruent with the focal advertisement message (i.e., the taste of the advertised product).

Practical Implications

The present findings have important implications for advertising and the design of multisensory marketing communications, since they provide a novel approach to communicating gustatory product attributes through crossmodal correspondences with the auditory modality. As the current study has shown that music can be reliably matched with basic tastes based on its auditory properties, it follows that these auditory properties carry meaning that may also serve as a nonverbal route to communicate taste properties. Specifically, consistent with general principles of the context-sensitive construal of meaning (for a review, see Schwarz, 2010), auditory properties would be decoded in terms of the relevant target dimension, that is, taste. To give a concrete example, marketers could include music designed to sound "sour" in an advertisement for a sour food product. Or, a candy store might want to play music having a "sweet" sound profile to crossmodally communicate the predominant taste of its products. Current marketing practice often uses music in food contexts without harnessing its full sound-symbolic potential. For example, ice cream manufacturer Häagen-Dazs's "concerto timer" app provides an augmented-reality music performance to indicate when the ice cream has reached its optimal consumption temperature. Scanning the barcode of a Häagen-Dazs ice cream container triggers the music, which is, however, identical across different ice cream flavours. By using the sound profiles identified in the present research, the soundtracks could easily have been engineered to create a consistent music-flavour experience.

This application of crossmodal correspondences may be an especially powerful tool in those contexts where primary product attributes are difficult or impossible to communicate (as in the case of advertising food products in audio-visual media), or where using verbal information (e.g., "extra sweet") is expected to trigger consumer reactance, or otherwise undesirable. Under such conditions, systematically designed music may be used together with other product features such as brand name, packaging shape, color, etc. to signal gustatory product attributes.

As the decoding of musically embedded information was observed across different cultural backgrounds in the present research (although with differing effectiveness), it seems likely that marketers can successfully use the auditory modality to communicate taste properties across different markets. However, while marketers targeting Western markets can use the musical parameters identified in the present research, those targeting non-Western markets should be aware that the proposed musical parameters might be less effective in their respective target markets. In the latter case, it may be necessary to further explore the extent to which the musical parameters identified in the current study elicit the intended taste associations in the specific target group. Additional research (applied and academic) may help to determine whether and how musical parameters can be optimally adjusted to the desired effect in specific non-Western markets. In sum, the present findings provide preliminary evidence that a global music strategy to communicate basic taste properties may be feasible, but that local adaptations may further increase its effectiveness.

Another question relevant to marketers who wish to apply the present findings is whether the acoustic profiles identified in the present study can be applied to different types of music (e.g., different musical genres). Would the effectiveness of the music-taste profiles vary as a function of musical genre? One argument seemingly contradicting this notion would be that different musical stimuli were used in the present research. While Experiment 1 used a short sequence of musical chords without further musical context, Experiments 2 and 3 used short pieces of ambient music including various acoustic and electronic instruments. The fact that sound-taste mappings were observed across these different musical contexts indicates that the effects are somewhat independent of specific musical forms or genres. In general, the identified parameters can be expected to be compatible with a musical genre as long as the variation of the parameters does not violate the genre's defining characteristics. It is important to note that the auditory parameters used in the present study are low-level, psychoacoustic parameters (rather than high-level parameters such as specific instruments or melodic motifs). While most musical genres are flexible enough to accommodate changes in low-level parameters such as pitch, roughness, or tempo, some genres may be incompatible with certain taste profiles (e.g., rap music and the high-pitched, legato profile of "sweet" music). In such cases, marketers are advised to switch to a more flexible genre.

It further seems likely that the processing of multisensory marketing stimuli by consumers can be enhanced if the crossmodal correspondences that exist between auditory cues and at least certain of the product or brand attributes match. Extant laboratory research in cognitive psychology and sensory neuroscience demonstrates enhanced multisensory integration of those sensory features that correspond crossmodally (see Parise & Spence, 2009; Spence, 2011, for a review). Enhanced multisensory integration, in turn, can result in (from a marketing perspective desirable) effects such as reduced sensory uncertainty (Munoz & Blumstein, 2012), increased attention capturing (Matusz & Eimer, 2011), and superadditive neural responses (Small, et al., 2004). Consistent with this idea, the beneficial effects of multisensory congruence on consumer preferences have been demonstrated for haptic-olfactory stimulus combinations and high-level semantic attributes (Krishna, Elder, & Caldara, 2010). Based on these findings from the multisensory integration literature, the simultaneous presentation of taste-congruent music and visual food product cues (e.g., images of food on product packagings or restaurant menus) may result in facilitated identification, increased visual saliency, and increased preference for the associated products.

Initial experimental evidence also suggests that expected or perceived properties of food products could be changed through the use of crossmodally congruent music (or vice versa). For instance, high- versus low-pitched background music has been shown to bias consumers' taste ratings of ambiguous bitter-sweet tastants (Crisinel, et al., 2012). Another recent, large sample study indicates that "sweet" versus "sour" background music can affect perceived taste attributes of wine (Spence, Velasco, & Knoeferle, in press), and ongoing work from our lab shows a trend, such that a bitter-sweet beverage was expected to be more bitter when advertised with a "bitter" soundtrack, and sweeter when advertised with a "sweet" soundtrack.[6] While it is not clear whether crossmodally congruent music influences low-level perceptual or rather decisional aspects of consumers' taste experiences, recent methodological innovations may help to disentangle these processes (e.g., Litt & Shiv, 2012).

References

Agarwal, P., Karnick, H., & Raj, B. (2013, 4-8 November 2013). *A comparative study of indian and western music forms.* Paper presented at the International Society for Music Information Retrieval Conference (ISMIR), Curitiba, Brazil.

An, D. (2007). Advertising visuals in global brands' local websites: A six-country comparison. *International Journal of Advertising: The Quarterly Review of Marketing Communications, 26*(3), 303-332.

Argo, J. J., Popa, M., & Smith, M. C. (2010). The sound of brands. *Journal of Marketing, 74*(4), 97-109.

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: Prevalence and familiality. *Perception, 25*, 1073-1079.

Bates, D. M., & Sarkar, D. (2013). lme4: Linear mixed-effects models using S4 classes, R package version 1.0-5.

Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality. *Psychological Science, 8*(4), 340-342.

Boesveldt, S., Frasnelli, J., Gordon, A. R., & Lundström, J. N. (2010). The fish is bad: Negative food odors elicit faster and more accurate reactions than other odors. *Biological Psychology, 84*(2), 313-317.

Coulter, K. S., & Coulter, R. A. (2010). Small sounds, big deals: Phonetic symbolism effects in pricing. *Journal of Consumer Research, 37*(2), 315-328.

Crisinel, A.-S., Cosser, S., King, S., Jones, R., Petrie, J., & Spence, C. (2012). A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background. *Food Quality and Preference, 24*(1), 201-204.

Crisinel, A.-S., & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience Letters, 464*(1), 39-42.

Crisinel, A.-S., & Spence, C. (2010a). As bitter as a trombone: Synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes. *Attention, Perception, & Psychophysics, 72*(7), 1994-2002.

Crisinel, A.-S., & Spence, C. (2010b). A sweet sound? Food names reveal implicit associations between taste and pitch. *Perception, 39*(3), 417-425.

Formatted: Spanish (Spain)

Formatted: French (France)

Crisinel, A.-S., & Spence, C. (2011). Crossmodal associations between flavoured milk solutions and musical notes. *Acta Psychologica, 138*(1), 155-161.

Crisinel, A.-S., & Spence, C. (2012a). A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses, 37*(2), 151-158.

Crisinel, A.-S., & Spence, C. (2012b). The impact of pleasantness ratings on crossmodal associations between food samples and musical notes. *Food Quality and Preference, 24*(1), 136-140.

Daltrozzo, J., & Schön, D. (2009a). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*(10), 1882-1892.

Daltrozzo, J., & Schön, D. (2009b). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Research, 1270*, 88-94.

Deroy, O., Crisinel, A.-S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review, 20*(5), 878-896.

Doyle, J. R., & Bottomley, P. A. (2011). Mixed messages in brand names: Separating the impacts of letter shape from sound symbolism. *Psychology & Marketing, 28*(7), 749-762.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114*(3), 405-422.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics* (3rd ed.). Berlin, Germany: Springer.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex, 20*(1), 109-120.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2012). Attention and multisensory feature integration. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 383-394). Cambridge, MA: MIT Press.

Fitzmaurice, G., Laird, N., & Ware, J. (2004). *Applied longitudinal analysis*. Hoboken, NJ: John Wiley & Sons.

Glendinning, J. I. (1994). Is the bitter rejection response always adaptive? *Physiology & Behavior, 56*(6), 1217-1227.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Holt-Hansen, K. (1968). Taste and pitch. *Perceptual and Motor Skills, 27*(1), 59-68.

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters, 11*(1), 5-20.

Klink, R. R., & Athaide, G. (2012). Creating brand personality with brand names. *Marketing Letters, 23*(1), 109-117.

Knöferle, K. M., & Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychonomic Bulletin & Review, 19*(6), 992-1006.

Kobayashi, M., Takeda, M., Hattori, N., Fukunaga, M., Sasabe, T., Inoue, N., et al. (2004). Functional imaging of gustatory perception and imagery: “Top-down” processing of gustatory signals. *NeuroImage, 23*(4), 1271-1282.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302-307.

Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: Liveright.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*(9), 635-642.

Krishna, A., Elder, R. S., & Caldara, C. (2010). Feminine to smell but masculine to touch? Multisensory congruence and its effect on the aesthetic experience. *Journal of Consumer Psychology, 20*(4), 410-418.

Kuehnl, C., & Mantau, A. (2013). Same sound, same preference? Investigating sound symbolism effects in international brand names. *International Journal of Research in Marketing, 30*(4), 417-420.

Litt, A., & Shiv, B. (2012). Manipulating basic taste perception to explore how product information affects experience. *Journal of Consumer Psychology, 22*(1), 55-66.

Lowrey, T. M., & Shrum, L. J. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research, 34*(3), 406-414.

Formatted: German (Germany)

Formatted: Swedish (Sweden)

Luna, D., Carnevale, M., & Lerman, D. (2013). Does brand spelling influence memory? The case of auditorily presented brand names. *Journal of Consumer Psychology, 23*(1), 36-48.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 384-394.

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review, 18*(5), 904-909.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243-247.

Mesz, B., Sigman, M., & Trevisan, M. (2012). A composition algorithm based on crossmodal taste-music correspondences. *Frontiers in Human Neuroscience, 6*(71), 1-6.

Mesz, B., Trevisan, M., & Sigman, M. (2011). The taste of music. *Perception, 40*(2), 209-219.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61-64.

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research, 178*, 67-77.

Moskowitz, H. R., Kluter, R. A., Westerling, J., & Jacobs, H. L. (1974). Sugar sweetness and pleasantness: Evidence for different psychological laws. *Science, 184*(4136), 583-585.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1273-1280.

Munoz, N. E., & Blumstein, D. T. (2012). Multisensory perception in uncertain environments. *Behavioral Ecology, 23*(3), 457-462.

Murray, N. M., & Murray, S. B. (1996). Music and lyrics in commercials: A cross-cultural comparison between commercials run in the Dominican Republic and in the United States. *Journal of Advertising, 25*(2), 51-63.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142.

O'Mahony, M., Goldenberg, M., Stedmon, J., & Alford, J. (1979). Confusion in the use of the taste adjectives 'sour' and 'bitter'. *Chemical Senses, 4*(4), 301-318.

Formatted: Swedish (Sweden)

Formatted: German (Germany)

Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology, 48*(5), 645-655.

Palmiero, M., Belardinelli, M. O., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A., et al. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing, 10*(2 Supplement), 268-271.

Parise, C. V., & Spence, C. (2009). “When birds of a feather flock together”: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *Plos One, 4*(5), e5664.

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception, 41*(7), 854-861.

Rudmin, F., & Cappelli, M. (1983). Tone-taste synesthesia: A replication. *Perceptual and Motor Skills, 56*(1), 118-118.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*(3), 225-239.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Schwarz, N. (2010). Meaning in context: Metacognitive experiences. In L. Mesquita, F. Barrett & E. R. Smith (Eds.), *The mind in context* (pp. 105-125). New York: Guilford.

Shrum, L. J., Lowrey, T. M., Luna, D., Lerman, D. B., & Liu, M. (2012). Sound symbolism effects across languages. *International Journal of Research in Marketing, 29*(3), 275–279.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception, 39*(4), 553-569.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., et al. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception, 35*(8), 1024-1033.

Small, D. M., Voss, J., Mak, Y. E., Simmons, K. B., Parrish, T., & Gitelman, D. (2004). Experience-dependent neural integration of taste and smell in the human brain. *Journal of Neurophysiology, 92*(3), 1892-1903.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971-995.

Formatted: French (France)

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology, 22*(1), 37-54.

Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception, 3*(5), 316-318.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition, 22*(1), 245-260.

Spence, C., Levitan, C. A., Shankar, M. U., & Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosensory Perception, 3*(1), 68-84.

Spence, C., Velasco, C., & Knoeferle, K. (in press). A large sample study on the influence of the multisensory environment on the wine drinking experience. *Flavour*.

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science, 4*(4), 653-667.

Tavassoli, N. T., & Lee, Y. H. (2003). The differential interaction of auditory and visual advertising elements with chinese and english. *Journal of Marketing Research, 40*(4), 468-480.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica, 30*(4), 201-213.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics, 74*(8), 1792-1809.

Walker, P., & Walker, L. (2012). Size–brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, & Psychophysics, 74*(6), 1226-1240.

Ward, J. (2013). Synesthesia. *Annual Review of Psychology, 64*, 49-75.

Watt, R., & Quinn, S. (2007). Some robust higher-level percepts for music. *Perception, 36*(12), 1834-1848.

Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research, 31*(1), 43-51.

Zhou, S., Zhou, P., & Xue, F. (2005). Visual differences in u.S. And chinese television commercials. *Journal of Advertising, 34*(1), 111-119.

Zhu, R. U. I., & Meyers-Levy, J. (2005). Distinguishing between the meanings of music: When background music affects product perceptions. *Journal of Marketing Research, 42*(3), 333-345.

Footnotes

[1] The present study follows a line of research using gustatory imagery triggered by taste words instead of actual tastants to study auditory-gustatory mappings (Crisinel & Spence, 2009, 2010b; Mesz, et al., 2012; Mesz, et al., 2011). Mental imagery occurs when perceptual representations in memory are accessed in order to predictively simulate actual perceptual events or properties (Kosslyn, Ganis, & Thompson, 2001; Moulton & Kosslyn, 2009). Using mental imagery triggered by basic taste words can be considered a valid instantiation of basic tastes for two reasons: 1) Gustatory imagery has been shown to activate the gustatory cortex in similar ways as actual gustatory stimuli (Kobayashi, et al., 2004; Palmiero, et al., 2009), and 2) gustatory imagery may capture the concept of individual basic tastes more accurately than actual tastants, which are often mistakenly identified by participants as flavors (e.g., see Simner, et al., 2010).

[2] Synesthesia is a neurological condition in which stimulation in one sensory modality leads to automatic, involuntary sensations in the same or another sensory modality (Ward, 2013). Depending on the applied definition, the prevalence in the general population is estimated to lie somewhere between 1 in 2000 (Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996) and 1 in 23 (Simner, et al., 2006).

[3] Attack was manipulated by changing the onset time of all notes in the musical sequence from 4.2 msec through 99 msec. Discontinuity was manipulated by changing the decay duration of all notes from 870 msec through 100 msec. The pitch manipulation was accomplished by transposing the musical sequences in quart intervals from C3 (130.81 Hz) through C7 (2093 Hz). Roughness was influenced through the application of a tremolo effect with a constant modulation frequency of 70 Hz at varying modulation amplitudes (0% to 100%). Sharpness was changed by applying a frequency filter that attenuated or boosted frequencies below/above 1000 Hz from

+12/-12 dB through -12/+12 dB. Tempo was varied from 65 through 195 BPM (all musical stimuli are available for download at https://soundcloud.com/crossmodal/sets/tastemusic.

[4] Two additional studies with public audiences in the UK (N = 15 and N = 19) resulted in even better decoding performance, thus providing corroborating evidence for the robustness of the findings reported in Experiment 2.

[5] The discussion whether (or which) crossmodal correspondences are innate versus learned and to which extent crossmodal correspondences are automatic processes is still on-going (Parkinson, Kohler, Sievers, & Wheatley, 2012; Spence & Deroy, 2012, 2013).

[6] In an additional experiment, the sweet and bitter soundtracks developed for Experiment 2 were played in the soundtrack of an advert for a food product. In particular, a food product (iced coffee) was selected that one might expect to have an identifiable yet ambiguous basic taste (sweet and bitter). The participants (N = 245) watched a short commercial of the product with either “sweet” music, “bitter” music, or no music, and a textual reference to either “delicious bitter-sweet taste” (high bitter-sweet salience) or “delicious coffee taste” (low bitter-sweet salience) (3 × 2 between-subjects design). Then the participants evaluated how sweet, bitter, likeable, and expensive they expected the coffee to be. Across the low-salience groups, there were trends, but no significant effects of music condition on expected sweetness (M_{sweet} = 4.63, M_{bitter} = 4.23, $M_{control}$ = 4.30) or bitterness (M_{sweet} = 2.80, M_{bitter} = 3.46, $M_{control}$ = 2.97), even after controlling for general preferences for sweet and bitter foodstuffs. No trends or significant effects emerged across the high-salience groups.

Table 1

Auditory properties of the original composition and the four derived basic taste compositions used in Experiment 1

		Soundtrack			
Auditory property	Original	Bitter	Salty	Sweet	Sour
Pitch	Medium	Low	Low	High	High
Discontinuity	Legato	Legato	Staccato	Legato	Staccato
Roughness (80 Hz tremolo)	Low	High	High	Low	High
Sharpness (-15 dB applied above/below 1000 Hz)	Low	Low	Low	Low	High
Consonance	Medium, G major	Low, G major dim	High, G major	High, G major	Occasionally low, G major dim

Table 2

Matching results between soundtracks and tastes labels in Experiment 2

		Taste response (in %)			
		Bitter	Sour	Salty	Sweet
Soundtrack	Bitter	49.2	34.4	9.8	6.6
	Sour	19.7	34.4	21.3	24.6
	Salty	21.3	18.0	49.2	11.5
	Sweet	9.8	13.1	19.7	57.4

Note. Contingency table depicting the distribution of responses for matching specific compositions (rows) with specific taste words (columns). Values on the diagonal represent correct mappings.

Table 3

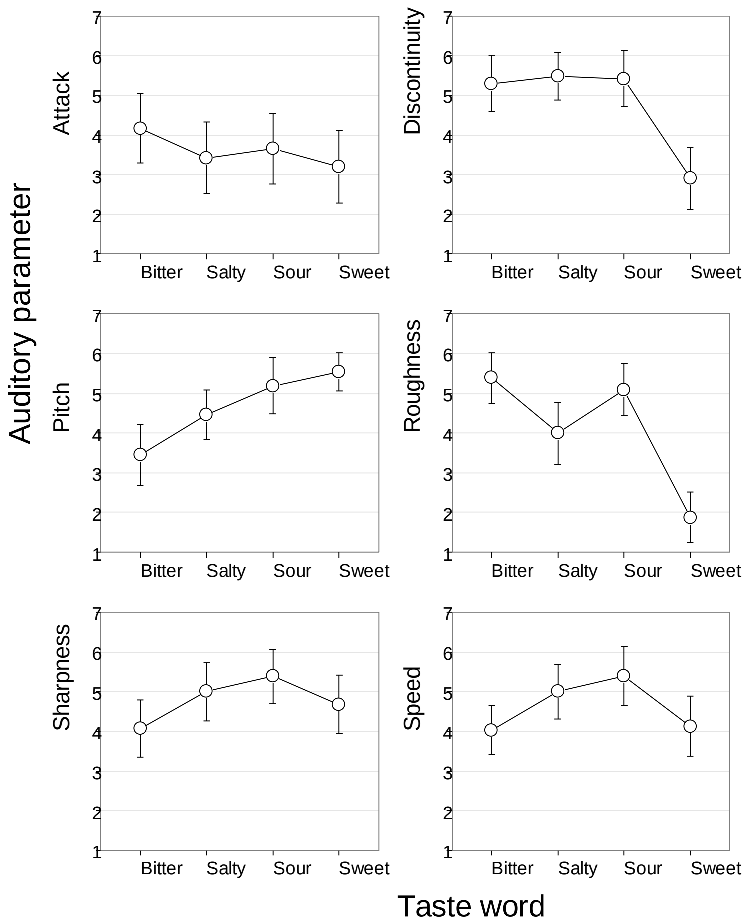
Comparison of models estimated for Experiment 3

	Model 1 (no country effect) correct ~ sound + musicYears + (1 | id)				**Model 2 (main effects model)** correct ~ country + sound + musicYears + (1 | id)				**Model 3 (interaction model)** correct ~ country * sound + musicYears + (1 | id)			
Log-Likelihood	-1277.72				-1273.78				-1271.69			
AIC	2567.45				2561.55				2563.38			
BIC	2601.17				2600.90				2619.59			
Marginal R^2	.068				.076				.078			
	b	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*
Intercept	-0.66	0.12	-5.67	.00	-0.89	0.14	-6.28	.00	-0.78	0.17	-4.48	.00
Sound (salty)	-0.27	0.14	-1.92	.05	-0.27	0.14	-1.92	.06	-0.30	0.22	-1.35	.18
Sound (sour)	-0.54	0.14	-3.78	.00	-0.54	0.14	-3.77	.00	-0.59	0.23	-2.58	.01
Sound (sweet)	0.78	0.14	5.72	.00	0.78	0.14	5.72	.00	0.46	0.21	2.16	.03
Music experience	0.04	0.01	4.14	.00	0.03	0.01	3.67	.00	0.03	0.01	3.68	.00
Country (US)					0.42	0.14	2.89	.00	0.24	0.22	1.07	.29
Country (US) × Sound (salty)									0.05	0.29	0.19	.85
Country (US) × Sound (sour)									0.09	0.29	0.31	.76
Country (US) × Sound (sweet)									0.54	0.28	1.93	.05

Note. Marginal R^2 describes the proportion of variance explained by the fixed factors (Nakagawa & Schielzeth, 2013).

Figure 1

Participants' selections for (psycho-)acoustic parameters in response to basic taste words in Experiment 1



Note. Error bars indicate 95% within-subjects confidence intervals (Morey, 2008).

Figure 2

Frequency distributions of correct responses in the observed data (bars) and under the assumption of purely random response behaviour (thick solid lines) in Experiment 2

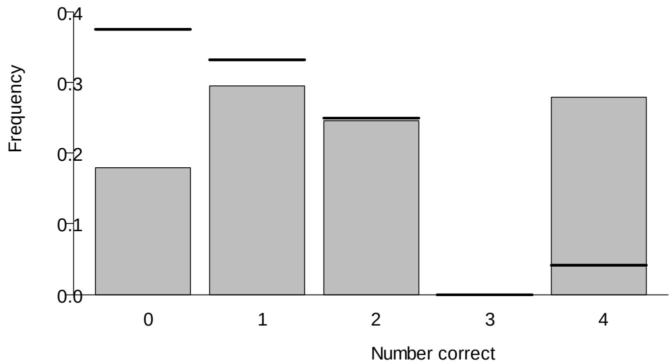
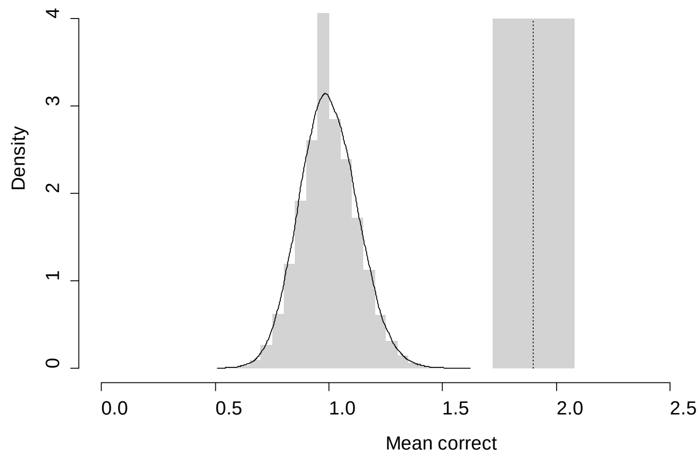


Figure 3

Comparison of bootstrapped versus observed data in Experiment 2



Note. Probability density function of average number of correct responses obtained from a Monte Carlo simulation under the assumption of random response behaviour (solid line, left) and mean (dotted line, right) and standard error of observed data.