

Crossmodal correspondences between sounds and tastes

Klemens Knöferle · Charles Spence

Published online: 6 October 2012
© Psychonomic Society, Inc. 2012

Abstract In this article, the rapidly growing body of research that has been published recently on the topic of crossmodal correspondences that involve auditory and gustatory/flavor stimuli is critically reviewed. The evidence demonstrates that people reliably match different tastes/flavors to auditory stimuli varying in both their psychoacoustic (e.g., pitch) and musical (e.g., timbre) properties. In order to stimulate further progress in this relatively young research field, the present article aims at consolidating prior findings concerning specific auditory–gustatory mappings, whereby special attention is given to highlighting (1) any conflicts in the existing experimental evidence and (2) any potential caveats with regard to the most appropriate interpretation of prior studies. Next, potential mechanisms underlying auditory–gustatory crossmodal correspondences are discussed. Finally, a number of potentially fruitful avenues for future research are outlined.

Keywords Crossmodal correspondences · Flavor · Music · Sound · Taste

Introduction

People often map stimulus properties from different sensory modalities onto each other in a manner that is surprisingly consistent. For instance, it has been demonstrated that there is robust agreement between people that rounded forms are associated with soft consonants (nasals and liquids such as [m] and [l]) and back vowels (such as [u]), whereas angular forms are more linked to hard consonants (plosives such as [t] and [k]) and front vowels (such as [i]) instead (Hinton,

Nichols, & Ohala, 2006; Köhler, 1929, 1947; Nuckolls, 1999). Such “crossmodal correspondences” have been described as consistent mappings between attributes or dimensions of stimuli (i.e., objects or events) either physically present (or else merely imagined) in different sensory modalities, be they redundant (i.e., referring to the same stimulus attribute) or not (Spence, 2011a).¹ As has been pointed out by Deroy and Spence (2012; Spence, 2011a), such nonarbitrary correspondences should be differentiated from various forms of crossmodal synaesthesia, which involve the conscious experience of concurrent stimuli not experienced by most people and are characterized by crossmodal mappings that are, among other distinguishing features, much more idiosyncratic in nature (Beeli, Esslen, & Jäncke, 2005; Martino & Marks, 2001). Crossmodal correspondences also need to be distinguished from two other forms of congruency between unisensory stimulus features—namely, spatiotemporal and semantic congruency—that have been studied in the context of the crossmodal binding problem (i.e., the question of how the brain determines which of the unisensory information available in the environment belongs to one and the same object; Spence, 2011a, 2012b). While spatiotemporal congruency refers to the proximity between two unisensory events in time and space (see, e.g., Spence & Driver, 2004) and semantic congruency refers to a crossmodal match (vs. mismatch) in terms of the identity or meaning of the unisensory component stimuli (see, e.g., Chen & Spence, 2010; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004), crossmodal correspondences arise between more basic, low-level stimulus features (Spence, 2011a).

¹ Although note that many other terms have been used, such as “cross-modal similarities” (Marks, 1987), “synaesthetic correspondences” (Melara & O'Brien, 1987), “synaesthetic associations” (Parise & Spence, 2008); for a detailed discussion of terminology, see Spence (2011a).

K. Knöferle (✉) · C. Spence
Department of Experimental Psychology, University of Oxford,
Oxford, UK
e-mail: klemens.knoferle@psy.ox.ac.uk

In the present article, we discuss the rapidly growing body of empirical literature that has started to investigate cross-modal correspondences between audition and taste, involving both simple (e.g., basic tastes, pure tones) and more complex (e.g., flavors, music) auditory and gustatory stimuli. Unlike combinations of other sensory modalities, such as vision and audition, correspondences between audition and taste have only recently started to receive serious scholarly attention (although see Fónagy, 1963, 2001, for a solitary early exception). Over the last couple of years, however, there has been a rapid growth of research documenting the existence of a variety of intriguing crossmodal mappings (or correspondences) between auditory and gustatory/flavor stimuli (e.g., Bronner, 2012; Bronner, Bruhn, Hirt, & Piper, 2008; Crisinel et al., 2012; Crisinel & Spence, 2009, 2010a, b, 2011, 2012; Mesz, Sigman, & Trevisan, 2012; Mesz, Trevisan, & Sigman, 2011; Simner, Cuskley, & Kirby, 2010; Spence, 2011a).² As such, we would argue that the topic is particularly appropriate and timely for review here.

Apart from the sheer number of recent publications, the present review is also motivated by several theoretical challenges that arise from previously reported findings on auditory–gustatory correspondences. First, a closer scrutiny of the empirical literature reveals, in some cases, conflicting results that require further discussion. Second, certain of the crossmodal mappings between sound and taste seem to defy previously proposed explanations of the underlying mechanisms of crossmodal correspondences. Third, the reported findings raise the question of why crossmodal correspondences between audition and gustation even exist in the first place. While crossmodal correspondences between, for instance, color and taste, appear to fulfil an apparent evolutionary purpose (e.g., red color and sweet taste are both cues for high nutritional value in fruit; Hoegg & Alba, 2007; Maga, 1974; Spence, Levitan, Shankar, & Zampini, 2010), the reason for the existence of auditory–gustatory mappings has not received a satisfactory explanation to date. The aim of the present review is thus to stimulate further research and theorizing in this field and to contribute to our understanding of how our senses interact, one with another.

Being a topic of potential practical relevance as well, this research area is now starting to attract the interest of those working in the food sector—not to mention marketers and advertisers interested in how they can convey something of the taste/flavor of their products simply by stimulating the ears of their potential customers ("Synaesthesia: Smells like Beethoven," 2012). The hope for many in the area is that our growing knowledge concerning the existence of crossmodal

correspondences may, for example, allow chefs, sommeliers, and/or food marketers to enhance their customers' dining experience by combining specific foods with crossmodally congruent sounds and music or to create acoustic stimuli that rely on crossmodal correspondences to highlight, or to set up (possibly subconscious), specific gustatory expectations concerning a given product (Crisinel, Cosser, King, Jones, Petrie & Spence 2012; Spence, 2011c; Spence & Shankar, 2010).

This article starts with a summary of prior findings regarding the existence of crossmodal correspondences between audition and taste. Special consideration is given to critically evaluating previous studies in terms of highlighting possible alternative explanations for the various results that have been obtained to date and identifying conflicting findings. In the section that follows, several plausible mechanisms that may underlie the reported crossmodal correspondences between audition and taste are discussed. Finally, the article concludes with a discussion of several potentially fruitful avenues for future research.

Summary of previous findings concerning crossmodal correspondences between gustation and audition

The basic idea that music may be described in terms of gustatory stimulus qualities is certainly not new. Indeed, one term that is often used to describe a style of musical expression is "*dolce*." This term is used to describe a gentle, literally "sweet" style of musical phrasing (Fallows, s.a.). Beyond such general analogies, though, it is worth noting that some composers have subjectively associated specific instruments with particular tastes; Berlioz, for instance, once described the "small acid-sweet" sound of the oboe (Berlioz, 1856; quoted in Mesz et al., 2011, p. 209; see also Baudelaire, 1857). In Huysmans's *À rebours*, the sensualist character Des Esseintes builds a "liqueur flavours keyboard" that allows him to play "on his palate a series of sensations analogous to those wherewith music gratifies the ear." For Des Esseintes, the taste of dry curaçao is like the "clarinet with its shrill, velvety note," while kümmel (liqueur) corresponds to "the oboe, whose timbre is sonorous and nasal" (Huysmans, 1926; pp. 59–61). Furthermore, composers have also used other idiosyncratic means of matching music to tastes/flavors. So, for example, the German composer Carl Gottlieb Hering used the notes c – a – f – f – e – e as the main motive of his "coffee canon" for children (Johnson, s.a.). In this case, however, the musical score was linked to taste in a purely symbolic manner.

Crossmodal mappings between tastes/flavors and sounds

Apart from such subjective and idiosyncratic historical linking of sounds to tastes/flavors, researchers have recently

² The present article reviews research examining crossmodal correspondences between basic tastes and complex flavors, on the one hand, and pure tones and musical sounds, on the other. The inclusion of flavors implies that olfactory sensations are at least partly involved in the formation of sound–flavor correspondences (Spence et al., 2010).

started to investigate crossmodal correspondences between audition and taste empirically. A growing body of research now demonstrates that Western participants will reliably match specific acoustic and musical parameters with different tastes, flavors, and oral-somatosensory food-related experiences (e.g., Bronner, 2012; Bronner et al., 2008; Crisinel & Spence, 2010a, 2010b; Mesz et al., 2011; Simner et al., 2010).

In two early studies, Holt-Hansen (1968, 1976) had participants match the pitch of a pure tone to two different brands of beers. Surprisingly, the average pitch that participants selected to match the taste/flavor varied as a function of the drinks presented. Specifically, Carlsberg's Elephant beer was matched to an average frequency of 640–670 Hz, while regular Carlsberg was matched to a tone with a frequency of 510–520 Hz. Even more intriguingly, some of the participants reported very rich, unusual, and pleasant sensory experiences when the perceived pitch and taste were judged to be “*in harmony*”—a quasi-synaesthetic condition that the author assumed to apply to only a small proportion of the population.

That said, a subsequent attempt to replicate Holt-Hansen's (1968, 1976) results was only partially successful (Rudmin & Cappelli, 1983). Specifically, although both the original studies and the replication found differing “optimal” frequencies for different kinds of beer, participants in the latter case did not report any of the extraordinary experiences originally described by Holt-Hansen. Since sample sizes were very small in both Holt-Hansen's original studies ($N_1=6$, $N_2=9$) and Rudmin and Cappelli's replication ($N=10$), such results should, perhaps, be interpreted with caution.

More recently, Crisinel and Spence (2009, 2010b) used a simplified version of the Implicit Association Test in order to test the strength of any crossmodal associations that might exist between the pitch of musical sounds and basic taste sensations. Specifically, the authors studied the associations between high versus low pitch and sour versus bitter tastes (Crisinel & Spence, 2009), as well as sweet versus salty tastes (Crisinel & Spence, 2010b). Their results suggested that sweet and sour tastes are associated with higher-pitched sounds, whereas bitter tastes are associated with sounds having a lower pitch (Crisinel & Spence, 2009). It is, however, unclear whether the reported findings describe a mapping of *relative* or of *absolute* pitch. That is, are sweet tastes mapped to a specific frequency of sound, or merely to whichever sound happens to have a higher frequency than any other sound available in the relevant stimulus set? In fact, the majority of studies of crossmodal correspondences to date have demonstrated relative effects (e.g., for audio-visual correspondences; Chiou & Rich, 2012). That said, it should perhaps be noted that the participants in the later studies of Crisinel and Spence (2011, 2012) chose moderate responses on the pitch scales (as opposed to selecting the scale endpoints). This might suggest a mapping of absolute pitch.

Since Crisinel and Spence (2009, 2010b) used food names rather than actual tastants as stimuli, any matching of such (imagined) tastes may have been inherently confounded with the linguistic features present in the food names themselves (e.g., the roundedness vs. angularity of the graphemes; Simner et al., 2010). Similarly, although not proposed previously, specific phonetic qualities of the speech sounds inherent in the food names themselves may, through a process of subvocalization, have influenced participants' choice of pitch.³

Finally, given that loudness perception in humans is nonlinear across the audible frequency range (Robinson & Dadson, 1957; Suzuki & Takeshima, 2004), changing the pitch or fundamental frequency of a harmonic sound while keeping the volume constant necessarily results in sounds with unequal loudness (remember that loudness is the perceptual correlate of physical sound intensity/volume that accommodates the human ear's differential sensitivity to specific frequencies and sound levels). This implies that pitch and loudness were most likely somewhat confounded in Crisinel and Spence's (2009, 2010b) studies, since all sounds were reportedly presented at a volume of 70 dB. Had these authors used pure tones (instead of synthetic instrumental sounds with more complex frequency spectra), the lowest pitch (D2=73.4 Hz) would roughly equal a loudness of 40 phon or 1 sone, while the highest pitch (C6=1046.5 Hz) would roughly equal a loudness of 70 phon or 8 sone (Suzuki & Takeshima, 2004). This means the latter sound would, in theory, have been 8 times louder than the former (Stevens, 1936).

Crisinel and Spence (2010a) addressed the concern about using only imagined tastes in a follow-up study by replicating the aforementioned experiment, but this time using actual tastants instead of just the names of foodstuffs. Again, sweet and sour tastes were consistently mapped to high-pitched sounds, bitter tastes to low-pitched sounds, and salty tastes to sounds having a medium pitch. In addition, this study highlighted the existence of a number of crossmodal correspondences between basic tastes and the sounds of various musical instruments (i.e., sounds that differed only in terms of their timbre). For instance, Crisinel and Spence's (2010a) results showed that bitter and sour tastes were reliably mapped to trombone sounds (which were, in turn, evaluated as sounding rather unpleasant by participants), while sweet tastes were typically mapped to piano sounds (evaluated by participants as rather pleasant). While this is certainly an interesting

³ Taking this idea further, it may theoretically be possible that even when given a food or beverage to taste, participants might still conjure up the name of that food or beverage and rely on linguistic and/or phonetic features for the crossmodal mapping. To rule out such a potential verbal mediation in the future, researchers might consider using an articulatory suppression procedure (see, e.g., Stevenson & Oaten, 2008).

finding in its own right, the question naturally arises as to which (psycho-)acoustic properties underlie this crossmodal effect (or mapping). In psychoacoustic terms, the sound of a trombone and the sound of a piano differ considerably in terms of their timbre. This difference is mainly attributable to a variation in prominent timbral dimensions—that is, the statistical distribution of component frequencies in the sound signal (usually referred to as spectral centroid, sharpness, or brightness; Fastl & Zwicker, 2007) and attack/onset time. Specifically, prior research into the perception of musical timbre suggests that the spectral centroid of brass instrument sounds is higher than that of piano sounds and that the attack of piano sounds is shorter than that of brass instrument sounds (Caclin, McAdams, Smith, & Winsberg, 2005; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). One means of manipulating these timbral dimensions of musical sounds in a more systematic manner (as opposed to using completely different instruments that may differ in terms of many of their acoustic parameters) would be to use synthesized sounds with carefully controlled acoustical properties. In addition, cultural associations for specific instruments (e.g., the frequent use of trombones in “sombre” music) may have had an influence on Crisinel and Spence’s (2010a) findings.

In another recent study designed to assess any crossmodal correspondences that might exist between acoustic qualities and basic taste sensations, Simner and her colleagues (2010) extended Crisinel and Spence’s (2009, 2010a, 2010b) original research to vowel sounds. These authors demonstrated that distinct properties of vowels could be systematically linked to the four most commonly asserted basic tastes (sweet, sour, salty, and bitter) and two levels of taste intensity. Sweet tastes were consistently associated with low spectral balance, low first formant frequency ($F1$, inversely associated with vowel height), and low second formant frequency ($F2$, inversely associated with vowel backness). Thus, sweet taste was mapped to high-tongue-position and back vowels. The mean values of the other basic tastes regarding these auditory characteristics increased in a sequence from sweet to bitter to salty to sour tastes, with sour tastes being mapped to low-tongue-position, front vowels. In terms of their temporal characteristics, the sweet taste corresponded to more continuous, smooth, vowel sounds without interruptions than did both the bitter and sour tastes. Finally, more concentrated tastes were mapped to increased spectral balance, higher $F1$ and $F2$ (i.e., low-tongue-position, front vowels), and higher vowel discontinuity (i.e., more staccato-like sounds). However, several of the auditory dimensions examined in this study are likely to have been confounded with auditory loudness, since Simner and her colleagues equalized physical volume, but not the psychoacoustic loudness of their auditory stimuli.

Crossmodal mappings involving musical sounds and/or complex flavors

The studies discussed so far suggest that crossmodal correspondences exist between low-level (psycho-)acoustic properties (pitch, spectral balance, etc.) and basic tastes (sweet, sour, etc.). In further studies, to be discussed next, other researchers have focused on more complex auditory and gustatory stimuli, such as short musical sequences and flavors.

In an extension of their earlier studies described above, Crisinel and Spence (2011) also examined crossmodal correspondences between flavored milk solutions and instrument sounds. Consistent with their previous findings, these authors observed a significant effect of flavor on the choice of pitch (C2 to C6), and of instrument type (piano, strings, woodwinds, and brass). As in several of the studies discussed earlier, the experimental design did not allow the researchers to rule out possible confounds of loudness and both pitch and instrument type.

Bronner and his colleagues (Bronner, 2012; ; Bronner et al., 2008) looked for any crossmodal correspondences between music and flavors. In a first study, participants tasted vanilla- and citrus-flavored drinks and then adjusted several auditory properties to fit the flavor stimuli. Using descriptive analysis rather than inferential statistics, the authors identified differing auditory associations for vanilla and citrus: The vanilla flavor corresponded to a soft, dull timbre that was neither sharp nor rough, a small range ambitus, an even, legato articulation, a nonsyncopated rhythm, a melody with small step intervals, and a slow tempo. In contrast, the citrus flavor was matched to a timbre that was bright, sharp, and rough, a medium-to-high range ambitus, an accentuated, staccato articulation, a syncopated rhythm, a melody with medium-to-large step intervals, and a lively and fast tempo. On the basis of the results of this first study, both shorter and longer pieces of music were then created to optimally represent the flavors of orange, lemon, grapefruit, and vanilla. The results suggested that individuals are able to correctly match musical pieces to flavors. That said, although the authors claimed that their research highlighted a link between musical attributes and flavors, it is unclear whether their results might not actually be described as highlighting the musical parameters associated with sweetness and acidity instead (Spence, 2012a), since these are prominent attributes of the said flavors (see Stevenson & Boakes, 2004).

A recent study by Knöferle and Spence (2012) provides preliminary evidence that people systematically map a series of psychoacoustic and musical properties onto basic tastes. Using a slider controlling one auditory property of a short chord progression at a time (synthesized from pure tones), participants selected, on each trial, the sound that best

matched a given basic taste word. First, the findings of Knöferle and Spence's study replicated the previously identified mapping between sweet tastes and higher pitch and bitter tastes and lower-pitched sounds (Crisinel & Spence, 2009, 2010b). Sour and salty tastes ranged between sweet and bitter tastes in terms of their pitch height. Second, the results indicate that people consistently map auditory roughness (Terhardt, 1974) onto basic tastes. On average, participants selected the lowest roughness values for sweet taste words, significantly higher values for salty tastes, and again significantly higher values for sour and bitter tastes. Third, participants consistently mapped sounds with a higher spectral centroid onto sour tastes and sounds with a lower spectral centroid onto bitter taste. Fourth, sweet tastes were mapped onto sounds low in discontinuity, whereas sour, salty, and bitter tastes were mapped onto high-discontinuity sounds. Fifth, a significant difference also emerged for musical tempo, with sour tastes resulting in the highest average tempo and bitter taste linked to the lowest average tempo. Finally, attack/onset time of the sounds was not reliably linked to any of the basic tastes. 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 loudness. This was accomplished by computing averaged loudness values for each stimulus as described in *ISO532 B* and then adjusting all sounds to the smallest obtained loudness value. However, since again, taste words rather than actual tastants were used in this study, an influence of typographic or phonetic features of the taste words on participants' auditory selections cannot be ruled out.

In another recent experiment designed to investigate any crossmodal correspondences between basic tastes and music, professional musicians improvised freely on the theme of basic taste words (e.g., "salty"; Mesz et al., 2011). The musical features of the musicians' resulting improvisations were coded and statistically compared and, as it turned out, reliably differed across the basic taste words. For example, the term "salty" consistently led to musical improvisations featuring staccato phrasings, while the concept of a "bitter" taste led to improvisations that were dominated by low-pitched sounds and legato phrasings. In a second experiment, those without any musical expertise were able to reliably map the musical improvisations to the four basic tastes. Similarly, a third experiment demonstrated that musical pieces created on the basis of the identified taste profiles using a computer algorithm were also reliably mapped (Mesz et al., 2012). The combined results from these experiments imply that analogies between auditory and gustatory cues are bidirectional, such that auditory cues can be mapped onto gustatory qualities and vice versa. Yet again, the fact that in the experiments, Mesz and his colleagues (2011) used taste words rather than actual tastants limits the informative value of the findings. Given that

music is often characterized in emotional terms and that it can communicate human and, particularly, emotional qualities (Fritz et al., 2009; Watt & Quinn, 2007), it is quite possible that at least some of the supposed taste words used to guide the musical improvisations in the first experiment (e.g., "sweet" or "bitter") were not interpreted exclusively as cues to basic tastes, but also as cues to emotional qualities (e.g., in English, people talk of feeling bitterness about . . .). Such an interpretation in emotional terms is even more likely to have happened since the musicians in Mesz et al.'s study were also asked to improvise on a number of nontaste, emotion words widely used in music (e.g., "feroce," "dolente"). What Mesz et al. would have shown, in the case of such an emotional interpretation, was a correspondence not between musical attributes and taste qualities, but between musical attributes and emotional concepts.

Focusing on flavors and speech sounds, Ngo, Misra, and Spence (2011) looked for any associations between chocolates varying in their cocoa content and "rounded" versus "angular" words. In three studies, the authors demonstrated that low-cocoa (i.e., sweet) chocolates are consistently associated with words featuring soft nasal and lateral consonants and back vowels (e.g., "lula," "maluma"), whereas high-cocoa (i.e., bitter) chocolates are associated with words containing hard plosive consonants and more front vowels (e.g., "tuki," "takete"). Although the chocolates used in the studies comprised complex flavors/foodstuffs, which are difficult to disentangle in terms of their most salient basic sensory qualities, the authors argued that the variation along the bitter-sweet continuum was the principal driver of the crossmodal correspondences that were observed. On a more critical note, one could argue as to whether the results provide any support for the existence of specifically auditory-gustatory correspondences at all. Indeed, since the verbal cues were always presented to participants visually, the crossmodal correspondences evidenced in this study may either partly, or even perhaps exclusively, have been driven by visual/typographic rather than auditory/phonetic factors of the stimuli. While one might think that such ambiguity could be resolved by presenting the stimuli auditorily, it is always going to be difficult to rule out completely the possibility that people do not simply imagine the visual form of the auditory concept (i.e., a graphemic representation).

Interim summary

The studies reviewed thus suggest that people are often able to map sounds to tastes in a manner that is non-random. On the basis of this previous research, the sonic elements that appear to be associated with each of the four most frequently mentioned so-called basic tastes are summarized in Table 1.

Table 1 Summary of crossmodal correspondences between basic tastes and sonic elements demonstrated to date

Author(s)	Auditory property	Sweet	Sour	Salty	Bitter
Bronner, 2012	Sharpness/spectral balance	low	high		
	Roughness	low	high		
	Ambitus	small	large		
	Articulation	legato	staccato		
	Rhythm	even	syncopated		
	Melodic intervals	small	large		
	Melodic consonance	consonant	dissonant		
	Tempo	slow	fast		
Crisinel & Spence, 2009	Pitch		high		low
Crisinel & Spence, 2010a	Pitch	high	high	average	low
	Instrument type	piano	brass	brass	brass
Crisinel & Spence, 2010b	Pitch	high	high	ns	ns
Crisinel & Spence, 2012	Pitch	higher			lower
	Instrument type	piano			ns
Knöferle & Spence, 2012	Pitch	high	average	average	low
	Roughness	low	high	average	high
	Sharpness/spectral balance	ns	high	ns	low
	Discontinuity	low	high	high	high
	Attack	ns	ns	ns	ns
	Speed	ns	fast	ns	slow
Mesz et al., 2011	Pitch	average	high	low	low
	Articulation	legato	average	staccato	legato
	Loudness	soft	average	average	average
	Chord consonance	consonant	dissonant	average	average
	Melody consonance	consonant	dissonant	average	average
Ngo et al., 2011 ²	Consonants	soft			hard
	Vowel backness	back			front
Simner et al., 2010	Vowel height	higher	lower	lower	lower
	Vowel backness ³	back	front	front	front
	Discontinuity	lower	higher	ns	higher
	Spectral balance	lower	higher	ns	ns

¹ One might argue that the observed difference in instrument choice in fact reflects differences in the acoustic parameters of these instrument sounds—namely, the spectral centroid (higher for brass) and the onset time (shorter for piano).

² In this study, basic tastes were not manipulated directly and independently, but only indirectly through chocolate with varying cocoa content.

³ These effects were significant but failed to survive Bonferroni correction.

By and large, the crossmodal correspondences reported so far show several striking consistencies. For example, the results of many studies now indicate that higher-pitched sounds are reliably mapped onto sweet and sour tastes, whereas lower-pitched sounds are mapped onto bitter tastes (Crisinel & Spence, 2009, 2010a, 2010b, 2012; Knöferle & Spence, 2012; Mesz et al., 2012; Mesz et al., 2011). Furthermore, in all of the studies described above that involved a manipulation of the spectral balance of sounds, sweet tastes were mapped onto lower spectral balance, while sour tastes were mapped onto higher spectral balance (Bronner, 2012; Knöferle & Spence, 2012; Simner et al., 2010).

Interestingly, Simner et al. (2010) noted in their study that sour tastes mapped onto higher values of $F1$ and $F2$ as well.

By integrating this finding with the results of Crisinel and Spence's (2009) study, which suggest that sour tastes map onto higher values of $F0$, Simner et al. arrived at the generalization that "it seems well supported that sour tastes may map to sound characteristics with higher frequency" (Simner et al., 2010, p. 563). While this observation would seem to be appropriate, it is important to note that the acoustic measures in question are linked to quite different perceptual correlates. Whereas fundamental frequency ($F0$) is *positively* correlated with the perceived pitch height of a sound (i.e., a higher $F0$ indicates a higher pitch), the formant frequencies $F1$ and $F2$ are *negatively* correlated to vowel height and vowel backness (note that higher $F1$ means lower vowels, while higher $F2$ means front vowels). In other words, an increase in $F0$ is

perceptually very different from an increase in $F1$, and the question therefore arises as to why these very different perceptual qualities should be similarly linked to sour tastes, or what underlying perceptual dimension would cause the effect in the first place. Unfortunately, Simner et al. (2010) did not vary vowels by pitch height ($F0$) in their experiment, and thus a direct comparison with the pitch-related results of Crisinel and Spence (2009, 2010b, 2012) is difficult to make.

To highlight another example of an apparently robust crossmodal mapping, in studies on correspondences between tastes and music, sweet tastes have consistently been associated with consonant chords and legato articulations, whereas sour tastes have been mapped to dissonant chords and staccato articulations (Bronner, 2012; Mesz et al., 2012; Mesz et al., 2011). On a more speculative note, such a crossmodal mapping may be attributable to perceived similarities in (hedonic) patterns of tension and relaxation. In music, dissonances have often been regarded as tensions, which are resolved by relaxations in the form of consonances (Bigand & Parncutt, 1999; Schenker, 1979). Similarly, sour tastes often elicit a tensing of the facial muscles (eye-squinting, nose-wrinkling) as part of an aversive reaction (Steiner, Glaser, Hawilo, & Berridge, 2001).

Findings concerning auditory–gustatory properties also exhibit consistencies with previous research on amodal stimulus properties. More specifically, previous research has found sour-tasting citrus flavor (Bronner, 2012) and the taste word “sour” (Knöferle & Spence, 2012) to be associated with fast musical tempos. Looking beyond the field of audition and taste, these results appear to be consistent with the observation that people naturally agree that lemons are fast rather than slow (B. C. Smith, 2012). Since tempo is an amodal stimulus property, one might therefore speculate that the mapping reported by these authors might actually be indicative of a more general correspondence between sour taste and fast tempo.

These consistent findings notwithstanding, it is worth pointing out that the present body of research also exhibits at least some inconsistency in terms of the crossmodal pairings that have now been documented. Whereas the results of Mesz et al.’s (2011) study highlighted a putative crossmodal correspondence between bitter tastes and the predominance of legato sounds, Simner et al. (2010) reported bitter tastes to be related to staccato vowel sounds instead. One potential explanation for this inconsistency would be that auditory–gustatory correspondences do not necessarily hold true for different classes of sounds (musical and speech sounds, respectively). However, this would seem rather unlikely, since the auditory feature in question—auditory discontinuity—is a low-level parameter that should be processed and perceived similarly in both domains. Alternatively, since Mesz et al. (2011) studied taste–sound mappings in a South American sample, it is always possible that cross-cultural differences may be responsible for the

conflicting results (e.g., Shankar, Levitan, & Spence, 2009). A comprehensive cross-cultural study including both musical and speech sounds could shed light on this question.

Despite such minor inconsistencies in the literature, other auditory characteristics exhibit comparable mappings between the domains of speech and nonspeech sounds. For instance, both Crisinel and Spence (2010a) and Simner et al. (2010) reported that sounds with low spectral balance (piano sounds, which are relatively low in high-frequency harmonics, and vowel sounds with low spectral balance, respectively) were associated with sweet tastes. Likewise, the sweet–legato and sour–staccato mappings appear to hold for both speech and musical sounds (Bronner, 2012; Mesz et al., 2011; Simner et al., 2010).

On a more general level, several studies now appear to challenge the view that the established basic tastes are theoretically separate and perceived as genuinely distinct sensations (Erickson, 2008). As was noted by Simner et al. (2010), this kind of distinct-sensation account would predict that each of the basic tastes should be mapped onto clearly distinct auditory parameters. Simner et al.’s own findings, however, as well as more recent results of Knöferle and Spence (2012), seem to suggest that not all of the basic tastes result in equally distinct mappings. For example, sour and salty tastes behave similarly when mapped onto pitch (Knöferle & Spence, 2012), vowel height, and vowel backness (Simner et al., 2010) and elicit responses that are clearly different from those seen for sweet taste. Similarly, sour, salty, and bitter tastes appear to group with each other when mapped onto auditory discontinuity (Knöferle & Spence, 2012) and instrument type (Crisinel & Spence, 2010a). As was proposed before, this finding might indicate a hierarchical ordering of basic tastes, in support of Erickson’s (2008) across-fibre patterning theory (Simner et al., 2010).

Possible mechanisms underlying crossmodal correspondences between tastes/flavors and auditory stimuli

Given that a multitude of crossmodal mappings has been observed in empirical studies by now, the question arises as to what mechanisms might explain the reliable (i.e., nonrandom) and seemingly ubiquitous matching of specific tastes to specific auditory stimuli. To date, several intriguing (albeit speculative) explanations have been put forward regarding the origins of the crossmodal correspondences between sound and taste/ flavor reviewed above.

Intensity matching

The first explanation regarding the origin of crossmodal correspondences is based on Stevens’s idea of intensity matching (Stevens, 1957; although see also earlier

discussions of sensory brightness: Cohen, 1934; Hornbostel, 1931; Schiller, 1935). Intensity matching describes the possibility of there being mappings between unimodal stimulus attributes that are magnitude based or *prothetic*—that is, that can be described in categories of “less” and “more.” In terms of auditory–gustatory/ flavor correspondences, one would, for example, expect auditory loudness to map onto the intensity of a gustatory stimulus, with an increase in one property being mapped onto an increase in the other property/dimension (L. B. Smith & Sera, 1992). Critically, if rigorous experimental control of loudness is not ensured in studies in this area, the matching of the intensity of sensations across different sensory modalities may easily be found or conceal other mappings.

Hedonic matching

An alternative idea is that people match tastes that are perceived to be unpleasant (e.g., bitter) with sounds that are less pleasant (e.g., trombone sounds), and more pleasant tastes (e.g., sweet) with more pleasant sounds (e.g., piano sounds; Crisinel & Spence, 2010a, 2012). In other words, certain crossmodal correspondences may be mediated by the common emotional valence of different stimuli. In general, such a hedonic matching account has received support from a variety of sources (see, e.g., Collier, 1996; Osgood, Suci, & Tannenbaum, 1957; Schifferstein & Tanudjaja, 2004). Relevant evidence pertaining to the case of crossmodal correspondences between audition and taste is, however, sparse. That said, the results of an experiment reported by Crisinel and Spence (2010a) showed that any matching of the pleasantness of different stimuli, if applicable to correspondences between gustatory and auditory stimuli, is by no means the whole story. These authors found that unpleasant tastants (e.g., very salty tastes) do not, in every instance, lead to a selection of sounds that are rated as unpleasant (i.e., the sound of a trombone). Even more strikingly, in another experiment designed to test for any crossmodal mappings between differently flavored chocolates (milk, marzipan, dark) and sounds varying in their pitch and instrument type, Crisinel and Spence (2012) reported that while their participants’ choice of instrument could be predicted by the pleasantness ratings they gave to the various chocolates, their choice of pitch was not. This finding clearly indicates that some other mechanism must underlie the taste/ flavor–pitch mapping.

Statistical co-occurrences

A third suggestion has been that the bitter–sweet to low–high vowel correspondence (Simner et al., 2010) may originate in the innate orofacial gestures that we all make in response to gustatory stimuli featuring basic tastes at birth (Spence,

2012a). That is, babies of different mammalian species, including humans, have been shown to protrude their tongue outward and upward in response to pleasant tastes and outward and downward in response to aversive tastes (Steiner et al., 2001). When air is exhaled, tongue positions such as these would be expected to result in high (e.g., [i] and [u], with lower $F1$) and low vowels (e.g., [a], with higher $F1$), respectively (Ladefoged & Johnson, 2011, pp. 22–23). This account can be thought of as a form of statistical co-occurrence that causes individuals to match naturally co-occurring unimodal stimulus attributes on the basis of their prior experiences. In this case, though, the statistical co-occurrence consists of the experience of an individual generating specific auditory cues in response to specific gustatory input early in life (or the observation of such behavior in other individuals). One might expect that such learned associations would be universal, if they are, for instance, related to physical laws (such as the physical mapping between the size of objects and the pitch and loudness of the sounds they make; see Parise & Spence, *in press*) or to experiences shared by all human (or mammalian) beings (Spence, 2012a).

Semantic matching

Fourth, semantically mediated correspondences may develop if the same terms or concepts are used to characterize sensations arriving from different sensory modalities. An example of the use of shared verbal labels would be the terms “high” and “low” in order to describe perceptual phenomena as different as auditory pitch, spatial configuration, and notes in a perfume. Along these lines, prior research has shown that people match high pitch with high spatial elevation, while matching low pitch with low spatial elevation (Melara & O’Brien, 1987; Pedley & Harper, 1959; Pratt, 1930; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006).⁴ One possible explanation for this ability to map sensory cues on the basis of common semantic features can be found in the semantic coding hypothesis (Martino & Marks, 1999, 2000). According to this account, these interactions between the senses may take place at a late stage of information processing—after information arriving in different senses has been encoded into a common, abstract (possibly semantic or even verbal) representation.

With regard to gustatory–auditory correspondences, the metaphorical use of taste words (e.g., “a sweet melody”) to describe auditory percepts has been addressed above. Regarding causality, it is hard to tell whether the consistent usage of a certain term (e.g., sensory metaphor; Marks,

⁴ At first glance, this particular mapping might appear to rather be an instance of intensity matching. Pitch height may seemingly reflect a prothetic stimulus attribute (i.e., one that differs in quantity), when, in fact, it is a metathetic dimension (i.e., one that differs in quality; cf. L. B. Smith & Sera, 1992; Spence, 2011a).

1991; Williams, 1976) across different modalities would be the cause or the consequence of a crossmodal mapping (see also Spence, 2011a). In other words, the question here is whether people map sweet tastes to what they perceive as sweet music because consistent terminology is coincidentally used to describe both modalities, or whether people use consistent terminology because of latent, nonsemantic relationships between the two senses. If the latter were found to be correct, the question would, of course, arise as to what kind of latent mechanism actually drives the consistent use of terminology (e.g., statistical co-occurrences or structural/functional properties of the brain).

Interim summary

To summarize, different accounts have been put forward, some (or all) of which may drive the auditory–gustatory crossmodal correspondences that have been observed to date. Note that these accounts should not necessarily be thought of as mutually exclusive alternatives but, rather, as possibly complementing each other—both in terms of explaining specific crossmodal correspondences and in terms of explaining crossmodal correspondences in general (Spence, 2011a). More specifically, there seems to be good reason to believe that different kinds of correspondences might be driven (or explained) by different mechanisms. The question, then, would be to ascribe specific mechanisms (or combinations of such mechanisms) to particular auditory–gustatory mappings, a venture that has only just begun to be tackled by researchers (see, e.g., Crisinel & Spence, 2012). Eventually, research in this area will hopefully lead to a better understanding of which of the crossmodal mappings between audition and taste stem from experiential or cultural factors and which, if any, are universal.

Speaking of universal correspondences, it should be noted that one specific mechanism known to drive other types of crossmodal correspondences—namely, structural correspondences—has not yet been tested in studies on auditory–gustatory correspondences. Structural correspondences are crossmodal mappings that originate from the specific organization or architecture of the cognitive system. Such correspondences may, for example, arise when two unimodal stimulus properties are represented by the same neural substrate or if certain neural connections are present at birth (Mondloch & Maurer, 2004). Identifying this type of mechanism as a driver for specific sound–taste mappings could, indeed, be seen as strong support for the universality of such mappings (Walsh, 2003).

Note that all of the mechanisms discussed above attempt to explain the regularities and patterns observed in crossmodal mappings. The question, however, as to why people should match specific tastes to specific sounds in the first place has not as yet been answered satisfactorily. While

crossmodal correspondences between, say, the visual and the gustatory modality can easily be hypothesized to serve an evolutionary purpose (e.g., both the redness and sweetness of a fruit can serve as indicators of its nutritional value; Hoegg & Alba, 2007; Maga, 1974; Spence et al., 2010), auditory–gustatory correspondences do not immediately seem to give rise to any obvious direct evolutionary advantage. One possible explanation for this phenomenon refers to the very nature of statistical learning, or predictive coding (Friston & Kiebel, 2009; Kilner, Friston, & Frith, 2007). Such a mechanism can be thought to be generally useful in that it will pick up on any regularities that are present in the environment, allowing individuals to make predictions on the basis of these regularities, and use the predictions to react more rapidly and accurately to environmental stimuli. In doing so, it may have originally evolved to facilitate the processing of fragmented unimodal information (Crisinel & Spence, 2011; Spence et al., 2010). However, such a general mechanism—as a by-product—will likely also pick up on correlations of sensory features that co-occur incidentally. For instance, the orofacial account proposed in the Statistical Co-Occurrences section is based on the idea that the experience of sweet (bitter) taste is subjectively linked to high-pitched (low-pitched) vocalizations, which may be a mere by-product of the facial expressions triggered by the respective taste. On the other hand, such a mapping may indeed serve an evolutionary purpose; it might have developed as a means for sharing information about the salubrity of food within social groups. In this case, an individual's high-pitched vocalizations in response to food stimuli would have signaled that the food was very likely high in nutrients and safe to ingest.

Avenues for future research

Identifying additional auditory parameters

The present state of research in the area of crossmodal correspondences provides ample opportunity for additional work. Future studies should, for instance, attempt to identify specific auditory parameters that correspond to particular taste qualities. For example, research in the field of musicology suggests that the mode (e.g., minor or major) of a scale, a chord, or even a song provides a strong indicator of its emotional valence, with the major mode being associated with positive valence and the minor mode being associated with negative valence (Gagnon & Peretz, 2003; Hevner, 1935). More recently, the argument has been put forward that the differential emotional effect of the minor and major modes may be related to acoustic patterns observable in animal vocalizations and, more particularly, in human speech (Bowling, Gill, Choi, Prinz, & Purves, 2009; Cook,

2007). Given that one source of crossmodal correspondences between auditory stimuli and tastes may be a hedonic matching of stimuli, it can be expected that the more pleasant mode (i.e., the major mode) will be matched to the more pleasant basic tastes (i.e., sweet; Moskowitz, Kluter, Westerling, & Jacobs, 1974), and flavors. In this context, it would also be informative to study the relationship between tastes/flavors and musical consonance. While in one, as yet unpublished study (Bronner, 2012), vanilla and citrus flavors were consistently mapped to music featuring high and low consonance, respectively, in another, the manipulation of consonance (i.e., consonant vs. dissonant triad chords) failed to give rise to a significant effect on the perceived taste of milk and dark chocolate (Spence & Shankar, 2010).⁵ The latter result seems surprising, given that consonance is a musical attribute that should be strongly related to hedonic valence and the hedonic value of background sounds has been shown to bias concurrent taste evaluations (Woods et al., 2011). In the light of these inconclusive results, we would suggest that it might thus be fruitful to revisit potential links between basic tastes and musical consonance, in terms of both matching basic tastes to consonant versus dissonant sounds and examining crossmodal effects of consonance on basic taste perception.

Examining crossmodal correspondences for additional gustatory parameters

In order to complement the auditory–gustatory correspondences that have been documented to date, it would also be worthwhile to examine which acoustic properties map onto the taste of umami (Kawamura & Kare, 1987). This may prove to be a particularly challenging venture in the West, since umami, despite being an indicator for protein-rich food (Yamaguchi & Ninomiya, 1999), is not a very familiar (or rather, identifiable) sensation to most people.

Opportunities for additional research also arise from studying crossmodal correspondences both in the domain of basic tastes and with regard to complex flavors. The question of whether to focus on basic tastes or complex flavors in crossmodal studies should be considered in the light of the ongoing discussion about whether or not basic tastes do, in fact, exist (Delwiche, 1996; Erickson, 2008). Given that complex flavors, rather than basic tastes, are what people normally experience in their daily lives, future research should perhaps pay increased attention to the existence of any

crossmodal correspondences regarding flavors. However, that said, studying flavors can be expected to be more challenging than studying basic tastes, since a generally agreed upon classification for flavors still does not exist.

Auditory–gustatory correspondences and multisensory processing

Comparing the nascent research on auditory–gustatory correspondences with the broader work on other crossmodal correspondences (particularly audio–visual) reveals important research questions and related experimental paradigms, which have not yet been addressed with regard to auditory–gustatory correspondences. In general, it is not yet clear how auditory–gustatory correspondences influence several key aspects of multisensory information processing.

For instance, it would be worthwhile to study whether correspondences between auditory and gustatory features affect selective attention and whether corresponding auditory–gustatory dimensions are integral or separable (Pomerantz & Garner, 1973; Shalev & Algom, 2000). Studies of other crossmodal correspondences have successfully used Stroop tasks and Garner’s speeded discrimination paradigm to examine these questions (Patching & Quinlan, 2002). While such evidence is still missing for audio–gustatory correspondences, previous research into Stroop interference between olfaction and gustation (White & Prescott, 2007), as well as interval–taste synaesthesia (Beeli et al., 2005), suggests that speeded reaction time designs can be used to study crossmodal correspondences between gustation and audition.

Similarly, regarding the stage of information processing at which auditory–gustatory correspondences might occur, it would be interesting to test whether these mappings are dependent on directing attention to the features involved. For this purpose, research on correspondences between other sensory modalities has used indirect tasks that draw attention away from the relevant features (e.g., Evans & Treisman, 2010). Such paradigms could readily be applied to study auditory–gustatory correspondences—for example, by varying auditory and gustatory features in a speeded classification task of another, unrelated auditory or gustatory property. Alternatively, temporal order judgment tasks could be used to examine this question (Kobayakawa & Gotow, 2011; Parise & Spence, 2009).

Furthermore, researchers could also potentially adapt the combined event-related potential and transcranial magnetic stimulation (TMS) methodology utilized recently by Bien, ten Oever, Goebel, and Sack (2012) in order to try and identify the neural substrates underlying auditory–gustatory correspondences (Spence & Parise, 2012). In their study, Bien and her colleagues used TMS in order to temporarily lesion the right intraparietal cortex—a brain region that is

⁵ Previous research has distinguished between two types of consonance: sensory consonance (the absence of annoying factors, also applicable to nonmusical sounds) and musical consonance (music-specific principles of tonal affinity, compatibility, and fundamental–note relation). In the present research, consonance refers to the latter, musical notion of the term, which includes sensory consonance (see, e.g., Fastl & Zwicker, 2007, pp. 362–366; Terhardt, 1984).

assumed to play a role in multisensory processing (Muggleton, Tsakanikos, Walsh, & Ward, 2007). Notably, the temporal lesioning eliminated the effect of crossmodal pitch–size congruency on multisensory integration (specifically, on the spatial ventriloquism effect) in the context of an auditory localization task. These findings suggest that the approach of Bien and colleagues may be applied to dissociate between different types of crossmodal correspondence.

The influence of individual and cultural differences on crossmodal correspondences

Another opportunity for research pertains to the role of individual differences in the perception of crossmodal correspondences. As was pointed out by Crisinel and Spence (2010b), one might expect musical expertise to moderate the strength and/or reliability of crossmodal correspondences between basic tastes and more subtle aspects of music, such as timbre or musical intervals. Therefore, the musical expertise of participants should be measured in future research. Similarly, gustatory expertise might act as a moderator (it should be noted, though, that participants' taste expertise did not have a significant effect on the strength of crossmodal effects or error rates in Crisinel & Spence, 2009).

Spence (2011c) has noted that cross-cultural investigations are likely to become more important, too (Henrich, Heine, & Norenzayan, 2010; see also Walker, 1987). Specifically, studying gustatory–auditory correspondences in non-Western populations may be instructive in terms of the better understanding that the results of such research might provide concerning the relative contributions of cultural–environmental versus phylogenetic influences on different types of crossmodal correspondences. Recent studies have started to make at least some progress along these lines, lending support to the claim of universality for specific types of crossmodal correspondences, while refuting it for others. To mention one particular example, intercultural studies of late indicate that African and Western participants share certain correspondences between shapes and speech sounds, while differing in terms of shape–taste correspondences (Bremner et al., 2012). Such results could certainly be taken to indicate that the tested shape–taste correspondences stem from culture-specific factors.

Effects of auditory–gustatory crossmodal correspondences on the perception of the unimodal component stimuli

The effects of crossmodal correspondences have traditionally been distinguished in terms of decisional and perceptual consequences (Spence, 2011a). While it seems very likely that stimuli sharing a crossmodal correspondence can have decisional effects, there have been tentative suggestions that such stimuli also have perceptual consequences (e.g., they

may sometimes influence the perception of the unimodal component stimuli).

However, empirical evidence of such perceptual effects of crossmodally congruent auditory and gustatory stimuli is, as yet, still rare. In the case of food products, a co-occurring auditory stimulus can enhance the perceived intensity of basic taste components that are crossmodally congruent with it (see also research into smell–taste congruence and enhancement; Schifferstein & Verlegh, 1996; Stevenson & Boakes, 2004). For example, soundtracks designed to be congruent with sweet tastes (i.e., containing higher-pitched sounds) intensify the perception of sweetness in a bittersweet toffee (Crisinel et al., 2012). The authors discuss two possible explanations for this effect. First, crossmodally corresponding stimulus attributes might give rise to enhanced multisensory integration, an explanation discarded by the authors as rather improbable given that the onset of the auditory cue preceded the onset of the gustatory cue. Second, crossmodally corresponding stimulus attributes might unconsciously stimulate expectancy effects by triggering specific expectations, thus priming the associated sensory attributes in the subsequent taste perceptions. As long as the actual experience is not too different from prior expectation, the final percept could then be expected to remain anchored to the initial expectation (see Schifferstein, 2001; Yeomans, Chambers, Blumenthal, & Blake, 2008).

Since the perception of taste and flavor appears to be particularly susceptible to the influence of nongustatory information (Elder & Krishna, 2010; Hoegg & Alba, 2007; Spence et al., 2010), congruency-related biases of taste perception may be more likely to occur than biases of a “more dominant” sensory modality (such as vision). Along these lines, Yorkston and Menon (2004) examined how brand names differing in vowel backness affected the expected creaminess of ice cream. Their results suggested that back vowel brand names (e.g., “Frosch”) led to higher expectations of product creaminess than did front vowel brand names (e.g., “Frisch”).

Another opportunity for research pertains to the directionality of auditory–gustatory crossmodal correspondences. While some research has been conducted to study the effects of stimuli featuring auditory–gustatory correspondences on gustatory perception, effects in the other direction have not yet, as far as we are aware, been examined. For example, it would be interesting to determine whether the gustatory properties of a food-stuff can influence the perception of simultaneously presented auditory stimuli (e.g., music).

While it thus seems very plausible that crossmodal congruency can influence perceptions of unimodal stimulus attributes, this effect likely depends on the particular modalities involved, on the nature of the correspondence (and the underlying mechanism), and on the task to be performed (Spence, 2011a).

Studying crossmodal correspondences in ecologically valid contexts

Related to the preceding section, a very promising direction for future research could be to transfer the findings concerning auditory–gustatory crossmodal correspondences to other relevant disciplines, such as consumer psychology and marketing. In the case of consumer psychology, for example, it has been found that products with semantically congruent tactile and olfactory features are evaluated more favorably than products with semantically incongruent features (Krishna, Elder, & Caldara, 2010; see also Seo & Hummel, 2011). This effect may be explained by findings demonstrating that semantic congruence can lead to enhanced behavioral performance (Chen & Spence, 2010; Laurienti et al., 2004), which, when metacognitively experienced as fluent processing, can, in turn, be expected to positively influence product evaluation (since fluent processing is known to be able to affect emotions and judgments; Reber, Winkielman, & Schwarz, 1998; Schwarz, 2004). Therefore, while semantic factors have received scholarly attention, the influence of crossmodal correspondences as a driver of multisensory congruence effects in evaluation and choice tasks has not yet been examined. In particular, it would be worthwhile to study the effects of crossmodal correspondences on consumers' preferences and behavior, both in the context of product perception and in the context of retail atmospherics. For example, would bitter products such as coffee benefit from being advertised using music that features crossmodally congruent acoustic properties, even if that involves musical features that are perceived as unpleasant? This congruency effect, should it emerge, would be particularly interesting, since it runs counter to conventional wisdom and might be interpreted building on feelings-as-information theory and mood misattribution (Schwarz & Clore, 1983, 2003). Overall, the opportunities for studying the marketing implications of crossmodal correspondences in the laboratory or in the field are manifold (see, e.g., Spence, 2011b, for a description of a collaboration with Starbucks) and are only now beginning to be understood (Nelson & Hitchon, 1995, 1999).

Relatedly, knowledge about crossmodal correspondences may also be applicable to the study of the various arts and, more generally, to the field of aesthetics. For instance, insights into crossmodal correspondences, especially those regarding their relationship to metaphors (Marks, 1982, 1991), may help to explain the aesthetic effects of specific artworks or even lead to a better understanding of art in general. In a more applied sense, the application of the findings reported here in art may allow the creation of works of art that lead to new aesthetic experiences.

Conclusions

As the findings reviewed in the present article demonstrate, there is now a solid body of experimental research to show that neurologically normal individuals map tastes (and other aspects of flavor/oral-somatosensation) and both musical and nonmusical sounds in a nonrandom manner. While findings like these further our understanding of the interaction of our various senses, they seem to be strikingly different from reports of “true” synaesthetes (otherwise referred to as cases of “strong” synaesthesia; Martino & Marks, 2001):

I've always had a connection between music and tastes in my mouth associated with particular instruments and notes. It is strongest when I listen to individual instruments and the clearer and less ‘muddy’ the pitch, the stronger the taste. Some examples: Violins taste like lemons. Cellos can be orange, or cherry if they play very low. Bass is cherry. Woodwinds tend to be ‘herbal’—like mint or some kind of herbal tea. (Day, 2011, pp. 18–19).

Thus, one important question in the field of multisensory perception still remains to be answered: Are “true” synaesthesia and crossmodal correspondences qualitatively different phenomena, or are they manifestations of a more or less continuous spectrum of crossmodal links, possibly rooted in the same neurocognitive mechanism (Beeli et al., 2005; Bien et al., 2012; Brang, Williams, & Ramachandran, 2012; Deroy & Spence, 2012; Gallace, Boschini, & Spence, 2011; Martino & Marks, 2001; Ward, Huckstep, & Tsakanikos, 2006)? Whatever the answer to this question turns out to be, the key point to note is that crossmodal correspondences between sounds and tastes/flavors reflect a robust empirical phenomenon with potentially widespread applications.

References

- Baudelaire, C. (1857). *Les fleurs du mal [Flowers of evil]*. Paris: Poulet-Malassis.
- Beeli, G., Esslen, M., & Jäncke, L. (2005). Synaesthesia: When coloured sounds taste sweet. *Nature*, 434(7029), 38–38.
- Berlioz, H. (1856). *A treatise upon modern instrumentation and orchestration, dedicated to Frederick William IV, King of Prussia (Vol. 7)*. London: J. Alfred Novello.
- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *Neuroimage*, 59(1), 663–672.
- Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, 62(4), 237–254.
- Bowling, D. L., Gill, K., Choi, J. D., Prinz, J., & Purves, D. (2009). Major and minor music compared to excited and subdued speech. *Journal of the Acoustical Society of America*, 127(1), 491–503.
- Brang, D., Williams, L. E., & Ramachandran, V. S. (2012). Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities. *Cortex*, 48(5), 630–637.

- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K., & Spence, C. (2012). “Bouba” and “Kiki” in Namibia? Western shape-symbolism does not extend to taste in a remote population. Manuscript submitted for publication.
- Bronner, K. (2012). What is the sound of citrus? Research on the correspondences between the perception of sound and taste/flavour. Manuscript submitted for publication.
- Bronner, K., Bruhn, H., Hirt, R., & Piper, D. (2008, 15–19 July). *Research on the interaction between the perception of music and flavour*: Paper presented at the 9th Annual Meeting of the International Multisensory Research Forum (IMRF), Hamburg, Germany.
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*(1), 471–482.
- Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404.
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339–353.
- Cohen, N. E. (1934). Equivalence of brightnesses across modalities. *The American Journal of Psychology*, *46*(1), 117–119.
- Collier, G. L. (1996). Affective synesthesia: Extracting emotion space from simple perceptual stimuli. *Motivation and Emotion*, *20*(1), 1–32.
- Cook, N. D. (2007). The sound symbolism of major and minor harmonies. *Music Perception*, *24*(3), 315–319.
- 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.
- 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. (2012). The impact of pleasantness ratings on crossmodal associations between food samples and musical notes. *Food Quality and Preference*, *24*(1), 136–140.
- Day, S. A. (2011). The human sensoria and synaesthetic approach to cooking. *Collapse*, *VII*, 378–409.
- Delwiche, J. (1996). Are there “basic” tastes? *Trends in Food Science & Technology*, *7*(12), 411–415.
- Deroy, O., & Spence, C. (2012). Synaesthesia reclassified: Borderline cases of crossmodally-induced experiences. Manuscript submitted for publication.
- Elder, R. S., & Krishna, A. (2010). The effects of advertising copy on sensory thoughts and perceived taste. *Journal of Consumer Research*, *36*(5), 748–756.
- Erickson, R. P. (2008). A study of the science of taste: On the origins and influence of the core ideas. *Behavioral and Brain Sciences*, *31*(1), 59–75.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 1–12.
- Fallows, D. (s.a.). “Dolce (i)”. *Grove Music Online. Oxford Music Online*. Retrieved from <http://www.oxfordmusiconline.com/subscriber/article/grove/music/07936>
- Fastl, H., & Zwicker, E. (2007). *Psychoacoustics* (3rd ed.). Berlin: Springer.
- Fónagy, I. (1963). *Die Metaphern in der Phonetik [The metaphors in phonetics]*. The Hague: Mouton.
- Fónagy, I. (2001). A research instrument. In I. Fónagy (Ed.), *Languages within language: An evolutive approach* (pp. 337–357). Amsterdam: John Benjamins.
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573–576.
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25–40.
- Gallace, A., Boschini, E., & Spence, C. (2011). On the taste of “Bouba” and “Kiki”: An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience*, *2*(1), 34–46.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*(2–3), 61–83.
- Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*(1), 103–118.
- Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (2006). *Sound symbolism*. Cambridge: Cambridge University Press.
- Hoegg, J., & Alba, J. W. (2007). Taste perception: More than meets the tongue. *Journal of Consumer Research*, *33*(4), 490–498.
- Holt-Hansen, K. (1968). Taste and pitch. *Perceptual and Motor Skills*, *27*(1), 59–68.
- Holt-Hansen, K. (1976). Extraordinary experiences during cross-modal perception. *Perceptual and Motor Skills*, *43*(3), 1023–1027.
- Hornbostel, E. M. (1931). Über Geruchshelligkeit [On smell brightness]. *Pflügers Archiv European Journal of Physiology*, *227*(1), 517–538.
- Huysmans, J.-K. (1926). *À rebours [Against the grain]*. London: Fortune Press (Original work published 1884).
- Johnson, D. M. (s.a.). “Round”. *Grove Music Online. Oxford Music Online*. Retrieved from <http://www.oxfordmusiconline.com/subscriber/article/grove/music/23960>
- Kawamura, Y., & Kare, M. R. (1987). *Umani: A basic taste: Physiology, biochemistry, nutrition, food science*. New York: Marcel Dekker.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166.
- Knöferle, K., & Spence, C. (2012). Crossmodal correspondences between basic tastes and auditory parameters. Manuscript submitted for publication.
- Kobayakawa, T., & Gotow, N. (2011). Interaction between olfaction and gustation by using synchrony perception task. *i-Perception*, *2*(8), 964–964.
- Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.
- Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.
- 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.
- Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston: Wadsworth Pub Co.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414.

- Maga, J. A. (1974). Influence of color on taste thresholds. *Chemical Senses*, *1*(1), 115–119.
- Marks, L. E. (1982). Synesthetic perception and poetic metaphor. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 15–23.
- 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.
- Marks, L. E. (1991). Metaphor and the unity of the senses. In H. T. Lawless & B. P. Klein (Eds.), *Sensory science: Theory and applications in foods* (pp. 185–205). New York: Marcel Dekker.
- Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*(7), 903–924.
- Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception*, *29*(6), 745–754.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*(2), 61–65.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323–336.
- 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.
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133–136.
- 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.
- Muggleton, N., Tsakanikos, E., Walsh, V., & Ward, J. (2007). Disruption of synaesthesia following TMS of the right posterior parietal cortex. *Neuropsychologia*, *45*(7), 1582–1585.
- Nelson, M. R., & Hitchon, J. C. (1995). Theory of synesthesia applied to persuasion in print advertising headlines. *Journalism and Mass Communication Quarterly*, *72*(2), 346–360.
- Nelson, M. R., & Hitchon, J. C. (1999). Loud tastes, colored fragrances, and scented sounds: How and when to mix the senses in persuasive communications. *Journalism and Mass Communication Quarterly*, *76*(2), 354–372.
- Ngo, M. K., Misra, R., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference*, *22*(6), 567–572.
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, *28*, 225–252.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Parise, C. V., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *442*(3), 257–261.
- 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.
- Parise, C. V., & Spence, C. (in press). Audiovisual crossmodal correspondences. In J. Simner & E. Hubbard (Eds.), *Oxford handbook of synaesthesia*. Oxford, UK: Oxford University Press.
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 755–775.
- Pedley, P. E., & Harper, R. S. (1959). Pitch and the vertical localization of sound. *The American Journal of Psychology*, *72*(3), 447–449.
- Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Attention, Perception, & Psychophysics*, *14*(3), 565–569.
- Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285.
- Reber, R., Winkelman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48.
- Robinson, D. W., & Dadson, R. S. (1957). Threshold of hearing and equal-loudness relations for pure tones, and the loudness function. *Journal of the Acoustical Society of America*, *29*(12), 1284–1288.
- Rudmin, F., & Cappelli, M. (1983). Tone-taste synesthesia: A replication. *Perceptual and Motor Skills*, *56*(1), 118.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*(2), 113–129.
- Schenker, H. (1979). *Free composition (E. Oster, Trans.)*. New York: Longman (Original work published 1935).
- Schiffstein, H. N. J. (2001). Effects of product beliefs on product perception and liking. In L. J. Frewer, E. Risvik, & H. N. J. Schiffstein (Eds.), *Food, people and society. a european perspective of consumers' food choices* (pp. 73–96). Berlin: Springer.
- Schiffstein, H. N. J., & Tanudjaja, I. (2004). Visualising fragrances through colours: The mediating role of emotions. *Perception*, *33*(10), 1249–1266.
- Schiffstein, H. N. J., & Verlegh, P. W. J. (1996). The role of congruency and pleasantness in odor-induced taste enhancement. *Acta Psychologica*, *94*(1), 87–105.
- Schiller, P. (1935). Interrelation of different senses in perception. *British Journal of Psychology: General Section*, *25*(4), 465–469.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality & Social Psychology*, *45*(3), 513–523.
- Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*(3/4), 296–303.
- Seo, H.-S., & Hummel, T. (2011). Auditory-olfactory integration: Congruent or pleasant sounds amplify odor pleasantness. *Chemical Senses*, *36*(3), 301–309.
- Shalev, L., & Algom, D. (2000). Stroop and Garner effects in and out of Posner's beam: Reconciling two conceptions of selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 997–1017.
- Shankar, M. U., Levitan, C. A., & Spence, C. (2009). Grape expectations: The role of cognitive influences in color-flavor interactions. *Consciousness and Cognition*, *19*(1), 380–390.
- Simner, J., Cuskey, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, *39*(4), 553–569.
- Smith, B. C. (2012). Lemons are fast. *Edge*. Retrieved February 27, 2012, from <http://www.edge.org/responses/what-is-your-favorite-deep-elegant-or-beautiful-explanation>
- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*(1), 99–142.
- Spence, C. (2011a). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995.
- Spence, C. (2011b). Sound design: Using brain science to enhance auditory and multisensory product and brand development. In K.

- Bronner, R. Hirt, & C. Ringe (Eds.), ((ABA))) *Audio Branding Academy Yearbook 2010/2011* (pp. 35–52). Baden-Baden: Nomos Publishers.
- Spence, C. (2011c). Wine and music. *The World of Fine Wine*, 31, 96–103.
- Spence, C. (2012a). 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. (2012b). Multisensory perception, cognition, and behavior: Evaluating the factors modulating multisensory integration. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 241–267). Cambridge: MIT Press.
- Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and cross-modal attention*. Oxford: Oxford University Press.
- 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., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. *i-Perception*, 3(7), 410–412.
- Spence, C., & Shankar, M. U. (2010). The influence of auditory cues on the perception of, and responses to, food and drink. *Journal of Sensory Studies*, 25(3), 406–430.
- Steiner, J. E., Glaser, D., Hawilo, M. E., & Berridge, K. C. (2001). Comparative expression of hedonic impact: Affective reactions to taste by human infants and other primates. *Neuroscience & Biobehavioral Reviews*, 25(1), 53–74.
- Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, 43(5), 405–416.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153.
- Stevenson, R. J., & Boakes, R. A. (2004). Sweet and sour smells: Learned synesthesia between the senses of taste and smell. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 69–83). Cambridge: MIT Press.
- Stevenson, R. J., & Oaten, M. (2008). The effect of appropriate and inappropriate stimulus color on odor discrimination. *Attention, Perception, & Psychophysics*, 70(4), 640–646.
- Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *Journal of the Acoustical Society of America*, 116(2), 918–933.
- Synaesthesia: Smells like Beethoven. (2012, February 4). *The Economist*, p. 66.
- Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica*, 30, 201–213.
- Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception*, 1(3), 276–295.
- Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Attention, Perception, & Psychophysics*, 42(5), 491–502.
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.
- Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2), 264–280.
- Watt, R., & Quinn, S. (2007). Some robust higher-level percepts for music. *Perception*, 36(12), 1834–1848.
- White, T. L., & Prescott, J. (2007). Chemosensory cross-modal stroop effects: Congruent odors facilitate taste identification. *Chemical Senses*, 32(4), 337–341.
- Williams, J. M. (1976). Synaesthetic adjectives: A possible law of semantic change. *Language*, 52(2), 461–478.
- Woods, A. T., Poliakoff, E., Lloyd, D. M., Kuenzel, J., Hodson, R., Gonda, H., et al. (2011). Effect of background noise on food perception. *Food Quality and Preference*, 22(1), 42–47.
- Yamaguchi, S., & Ninomiya, K. (1999). Umami and food palatability. In R. Teranishi, E. L. Wick, & I. Homstein (Eds.), *Flavor chemistry: Thirty years of progress* (pp. 423–432). New York: Kluwer Academic.
- Yeomans, M. R., Chambers, L., Blumenthal, H., & Blake, A. (2008). The role of expectancy in sensory and hedonic evaluation: The case of smoked salmon ice-cream. *Food Quality and Preference*, 19(6), 565–573.
- Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research*, 31(1), 43–51.