Article

PERCEPTION

Perception
2016, Vol. 45(10) 1099–1114
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0301006616652018
pec.sagepub.com
SAGE

Cross-Modal Correspondence Among Vision, Audition, and Touch in Natural Objects: An Investigation of the Perceptual Properties of Wood

Shoko Kanaya
Human Information Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Kenji Kariya
Tsukuba Research Institute, Sumitomo Forestry Company, Tsukuba, Japan

Waka Fujisaki
Human Information Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Abstract
Certain systematic relationships are often assumed between information conveyed from multiple sensory modalities; for instance, a small figure and a high pitch may be perceived as more harmonious. This phenomenon, termed cross-modal correspondence, may result from correlations between multi-sensory signals learned in daily experience of the natural environment. If so, we would observe cross-modal correspondences not only in the perception of artificial stimuli but also in perception of natural objects. To test this hypothesis, we reanalyzed data collected previously in our laboratory examining perceptions of the material properties of wood using vision, audition, and touch. We compared participant evaluations of three perceptual properties (surface brightness, sharpness of sound, and smoothness) of the wood blocks obtained separately via vision, audition, and touch. Significant positive correlations were identified for all properties in the audition–touch comparison, and for two of the three properties regarding in the vision–touch comparison. By contrast, no properties exhibited significant positive correlations in the vision–audition comparison. These results suggest that we learn correlations between multi-sensory signals through experience; however, the strength of this statistical learning is apparently dependent on the particular combination of sensory modalities involved.

Corresponding author:
Waka Fujisaki, Human Information Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan.
Email: w-fujisaki@aist.go.jp

Keywords
Cross-modal correspondence, Perceptual properties of wood, Vision, Audition, Touch, Natural objects

Introduction

In the human sensory system, a signal conveyed by one modality may cause a sensation in another modality with which it has no direct relationship, or we may find that there is a systematic relationship between information conveyed via two or more sensory modalities that are not directly related. Synaesthetes, for example, may "feel" a particular color when they look at a character written in black ink or hear a specific sound (Asano & Yokosawa, 2013; Ramachandran & Hubbard, 2001; Simner et al., 2006). Further, a systematic relationship may exist between objectively unrelated information between two sensory modalities that are experienced by the majority of people, including non-synaesthetes. For example, the combinations of a figure spatially positioned upwards and a sound with a high fundamental frequency, or a figure spatially positioned downwards and a sound with a low fundamental frequency, are often perceived as compatible (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Melara & O'Brien, 1987; Patching & Quinlan, 2002). This kind of compatibility has been reported to affect participants' performance in various psychological tasks such as the speeded discrimination/classification paradigm, the implicit association test (IAT), spatial localization, and temporal order judgment (Bernstein & Edelstein, 1971; Bien, ten Oever, Goebel, & Sack, 2012; Burr, Parise, & Spence, 2009; Evans & Treisman, 2010; Parise & Spence, 2008, 2012). This collection of phenomena is known as cross-modal correspondence, and is an important clue to the way information is integrated across different sensory modalities (see Spence, 2011, for review).

According to previous reports, correspondence exists among various sensory modalities, including vision, audition, touch, taste, and smell. Apart from the earlier-stated example of position versus pitch, a correspondence between vision and audition has been reported elsewhere, including for surface brightness/lightness and pitch (Marks, 1987; Martino & Marks, 1999; Melara, 1989), surface brightness/lightness and loudness (Marks, 1987; Smith & Sera, 1992), figure size and loudness (Gallace & Spence, 2006; Marks, Hammeal, Bornstein, & Smith, 1987; Mondloch & Maurer, 2004), figure shape and pitch (Marks, 1987), and figure shape and vocal sound type (Köhler, 1929; Ramachandran & Hubbard, 2001). As an example from another sensory modality, there have long been reports of a systematic relationship between colour and temperature (Bennett & Rey, 1972; Ho, Van Doorn, Kawabe, Watanabe, & Spence, 2014; Mogensen & English, 1926). In recent years, a new line of research has examined the correspondence of vision and audition with smell and taste (Crisinel, Jacquier, Deroy, & Spence, 2013; Crisinel & Spence, 2010; Deroy, Crisinel, & Spence, 2013; Simner, Cuskley, & Kirby, 2010).

It has been hypothesized that at least some of these cross-modal correspondences reflect inherent correlations between sensory signals mediated by multiple sensory modalities that are learnt by people in their natural environment (Parise & Spence, 2012). For example, as large objects inherently have a lower resonant frequency compared with small objects, a lower sound should normally be emitted from a larger object. Similarly, a high pitch and an angular shape are felt to be harmonious because a hard object has a higher resonant frequency and breaks into sharper pieces compared with a soft object. It remains unclear, however, whether cross-modal correspondence occurs because of inferences by the sensory system based on learnt cross-modal correlations. To answer this question, cross-modal correspondence needs to be investigated not only in relation to the artificially generated stimuli that previous studies have used but also in relation to natural objects. If the sensory system in fact extracts and learns correlations between multi-sensory signals, then in addition to directly perceiving a natural object's property (e.g., surface brightness) with a specific sensory modality (e.g., vision), the system should also be able to infer that property with reasonable reliability based on sensory signals in other modalities (e.g., audition or touch).

In line with this hypothesis, some previous studies have reported cross-modal correlations in perceiving natural objects. Klatzky, Pai, and Krotkov (2000) examined how the perceived material and length of objects may be reflected by sounds. Grassi (2005) examined whether the size of an object can be estimated by the sound of it dropping on another object. Both studies showed that acoustic features are correlated with estimations of object material, length, and size. A more recent report by Velasco, Jones, King, and Spence (2013) found that participants could tell whether water is hot or cold by listening to the sound of it pouring into receptacles. These findings may support the idea that we learn and utilize inherent correlations between sensory signals mediated by multiple modalities, through experience of objects around us in our daily lives. Nonetheless, there are still limited numbers of such reports, and they only address some perceptual properties of specific sensory modalities. To verify the hypothesis, large-scale data derived from an investigation of various perceptual properties of natural objects are needed.

Fujisaki, Tokita, and Kariya (2015) conducted an experiment investigating the perception of the material properties of wood (a natural substance), in which vision, audition, and touch were independently evaluated by participants. Twenty-two types of test pieces were prepared and 50 participants were asked to separately evaluate either (a) photographs that had been taken of the pieces of wood (the visual condition), (b) the sound of the wood being hit (the auditory condition), or (c) the way that the pieces felt to the touch (the tactile condition), by assigning ratings to 23 materially descriptive adjectives. Their dataset was expansive. Although Fujisaki et al. (2015) have reported results from separate analyses of participant evaluations of vision, audition, and touch, this dataset is also ideal for the investigation of cross-modal correspondence of natural objects. As all participants made evaluations of the test pieces using adjectives across all modality conditions, it is possible to examine the degree to which participant ratings using different modalities are related. Hence, the present study reanalyses part of the dataset of Fujisaki et al. (2015) to determine the extent to which participants' evaluations on perceptual properties of wood conform across different modalities.

If cross-modal correspondence results from the statistical learning of correlations between multi-sensory signals in the natural environment, participants' evaluations of a perceptual property of wood (a natural object) in a given modality will agree to a certain extent with evaluations of that object's property made using a different modality. That is, we would observe cross-modal correspondence for the perception of properties of wood. The present study therefore reanalyzed part of Fujisaki et al.'s dataset (2015) in order to determine the extent to which participants' evaluations are in line with the hypothesis of cross-modal correspondence regarding vision, audition, and touch. Specifically, we compared the perception of three material properties each unique to vision, audition, and touch (surface brightness, sharpness of sound, and smoothness), among vision, audition, and touch.

Method

The following is a brief description of the method; full details can be found in Fujisaki et al. (2015).

Participants and Stimuli

Participants were 50 adults aged 20–40 years old with normal or corrected-to-normal vision and hearing abilities. Stimuli were 22 pairs of test pieces, each pair was made of two blocks of a single type of wood (untreated, treated, or fake). The first of each pair comprised Series 1; the second comprised Series 2.[1] All test pieces were of uniform size: 60 mm width, 120 mm length, and 9 mm thick. The surfaces of all test pieces were sanded with 240-grit sandpaper. Of the 22 types, this study only addressed ratings of 14, which were all untreated samples of different tree species.[2]

Visual stimuli were photographs taken of each test piece (e.g., Figure 1(a)). All photographs were presented on a display with the same visual angle relative to the observer. Auditory stimuli were recordings of each test piece being struck with a mallet: each test piece was placed onto a xylophone from which all boards had been removed and the ball of a mallet was placed to fall onto the test piece from a height of 1 cm. All sounds were generated by striking the pieces in this fashion. The recorded sound was edited to a duration of 1 s and then repeated five times, so that the duration of each auditory stimulus was 5 s. See Figure 1(b) for an example of the waveform and spectrogram. Auditory stimuli were presented through headphones. Tactile stimuli were the original test pieces placed into a small paper case with an opening in the top panel. Participants inserted their index finger through the opening to touch the test piece. The paper case was placed into a box with a black curtain drawn around it, so that participants were prevented from seeing the test pieces. Figure 1(c) shows participants touching the tactile stimuli.

Procedure

In each trial, a single stimulus (a photograph, a sound recording of the test piece, or the test piece itself) was presented. Twenty-three pairs of opposite-meaning adjectives were presented on a display concurrently with the stimulus; participants subsequently assigned ratings to these using a seven-point scale (some of these adjectives were used in the analysis for this study). Twenty-two trials were included in a session. In each session, stimuli from Series 1 or Series 2 were presented in a single modality in random order. Two sessions were, therefore, held in the vision, audition, and touch conditions, for a total of six sessions overall. These sessions were presented in random order, although the first session was restricted to the vision condition. The first session was conducted in the visual domain because it was the most understandable condition for participants and helped them to get used to all the experimental procedures.

The adjective pairs ranged from adjectives related to relatively early perceptual processing to those related to relatively higher cognitive processing (for more details, refer to Fujisaki et al., 2015). Former adjective pairs included those referring to properties sensed by vision, audition, and touch. In particular, the adjective pairs referring to visual properties were ''matte surface/gloss surface'', ''dark surface/bright surface'', and ''dull surface/clear surface''. Adjective pairs that referred to auditory properties were ''dampened sound/ringing sound'', ''dull sound/sharp sound'', and ''mixed sound/pure sound''. Adjective pairs that referred to haptic properties were ''rough/smooth'', ''cold/warm'', ''soft/hard'', ''light/heavy'', ''dry/wet'', and ''sparse/dense''. It was difficult for participants to assign ratings via sensory modalities that are not typically involved in perceiving the relevant property. Participants were, therefore, instructed to make speculative ratings in such cases. For example, participants rating the surface brightness of a stimulus in the auditory condition made ratings by estimating how bright the surface was likely to be purely based on the sound.

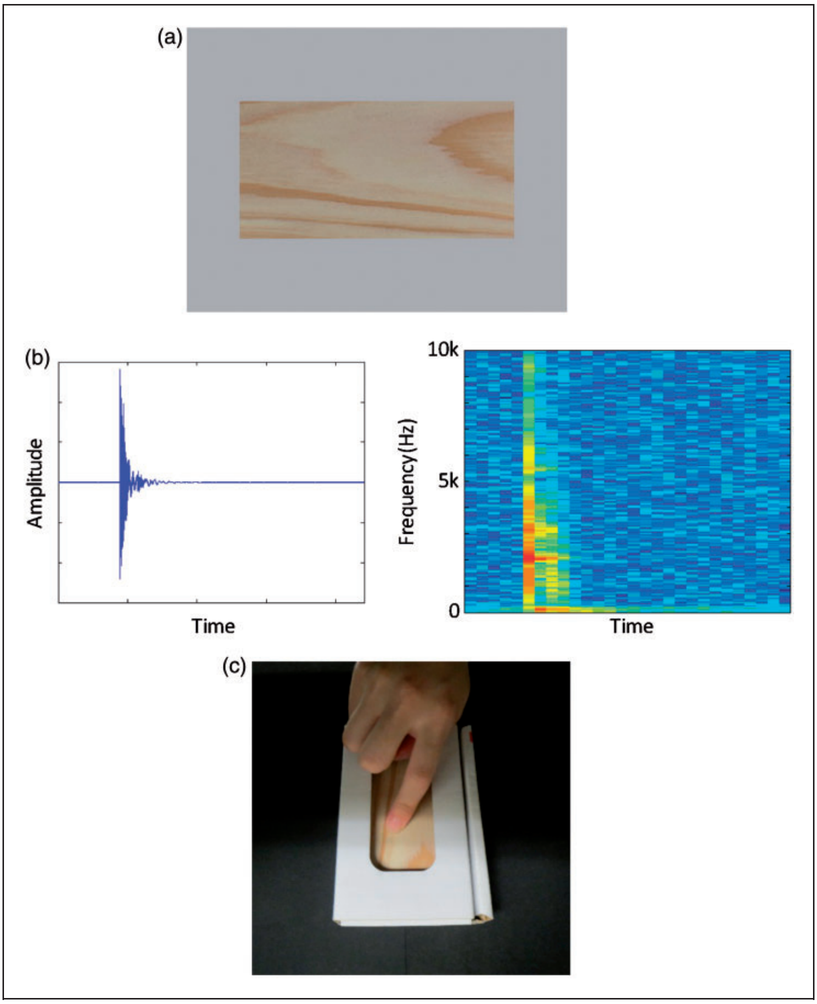

Figure 1. An example of a stimulus and the experimental conditions applied in Fujisaki et al. (2015). (a) A photograph of one of the wood blocks used as a visual stimulus. The photograph shown is of Japanese cedar (Series 2). (b) The sound of wood being struck used as an auditory stimulus. The waveform and spectrogram shown are of the sound of a piece of Japanese cedar (Series 2) being struck by a mallet. (c) The conditions of the tactile condition of the experiment.

This experiment was approved by the Institutional Review Board of the National Institute of Advanced Industrial Science and Technology (AIST), and was performed in accordance with the Declaration of Helsinki.

Adjectives Used in the Analysis

In the present study, we examined the degree participant ratings from different modalities corresponded, focusing on three perceptual properties: surface brightness, sharpness of the sound, and smoothness. We selected these three properties for the following reasons. In the following, we refer to a modality typically involved in perceiving a property and that may directly perceive it as a "direct" modality (e.g., vision for surface brightness). By contrast, if a modality is not typically involved in perceiving the property and cannot perceive it directly, this constitutes an ‘‘indirect’’ modality (e.g., audition for surface brightness). In testing, if ratings for a property based on modality A (a direct modality) are related to ratings based on modality B (an indirect modality), it is necessary to address a property for which participants' direct-modality ratings are highly consistent, or well correlated between participants. For example, to test if participants are able to correctly estimate smoothness from audition (an indirect modality relative to surface smoothness), it is necessary to know first that smoothness is at least consistently perceived by touch, in order to be able to make the comparison. We, therefore, addressed the three properties (one corresponding to each modality) that exhibited the highest inter-participant correlation coefficients (IPC) in direct rating in Fujisaki et al. (2015). High IPC values indicate small individual differences in ratings of the property, and that the participants' ratings of the property were consistent. The adjective pairs with the highest IPC values were ‘‘dark surface/bright surface’’ for the visual condition, ‘‘dull sound/sharp sound’’ for the auditory condition, and ‘‘smooth surface/rough surface’’ for the tactile condition.

Results

Figure 2 presents results for surface brightness; Figure 3 presents results for sound sharpness; Figure 4 presents results for smoothness.

Brightness of Surface

First, to determine that the degree of ‘‘surface brightness’’ judged by vision is consistent with ‘‘surface brightness’’ judged by audition and touch, we conducted a correlation analysis for the V–A, V–T, and A–T comparisons. The results are presented in Figure 2(a). We observed a significant positive correlation in the A–T comparison ($p < .01$) and a significant negative correlation in the V–A comparison ($p < .01$). If participants were able to correctly estimate surface brightness not only from vision but also from audition and touch, the correlation coefficients would have been significant and positive for the V–A and V–T comparisons. This was not the case, however; the correlation coefficient was negative for the V–A comparison. Participants were therefore unable to judge ‘‘surface brightness’’ using audition and touch, as they did when using vision. Nonetheless, a significant positive correlation was found between ratings made by audition and touch, despite the fact that participants made speculative estimations of surface brightness in these two modalities. This suggests that some commonalities were present between ratings of surface brightness in the two indirect modalities.

We subsequently conducted a further correlation analysis to determine if the relationship between ‘‘surface brightness’’ judged by vision is related to auditory properties such as ‘‘ringing of the sound’’, ‘‘sharpness of the sound’’, and ‘‘pureness of the sound’’ as judged by audition. Correlation coefficients were calculated for three comparisons between ‘‘surface brightness’’ and each of these auditory properties. The results are presented in Figure 2(b). Significant negative correlations were found for all comparisons ($p < .01$). These results suggest that wooden pieces with bright surfaces were likely to be judged as making dampened, dull, and mixed sounds.

The aforementioned correlation analysis examined co-variations between the two variables, but is unable to justify inferences of causal relationships between them. We, therefore, conducted a multiple regression analysis to identify auditory properties able to predict ‘‘surface brightness’’ as judged by vision. In this analysis, ‘‘surface brightness’’ judged

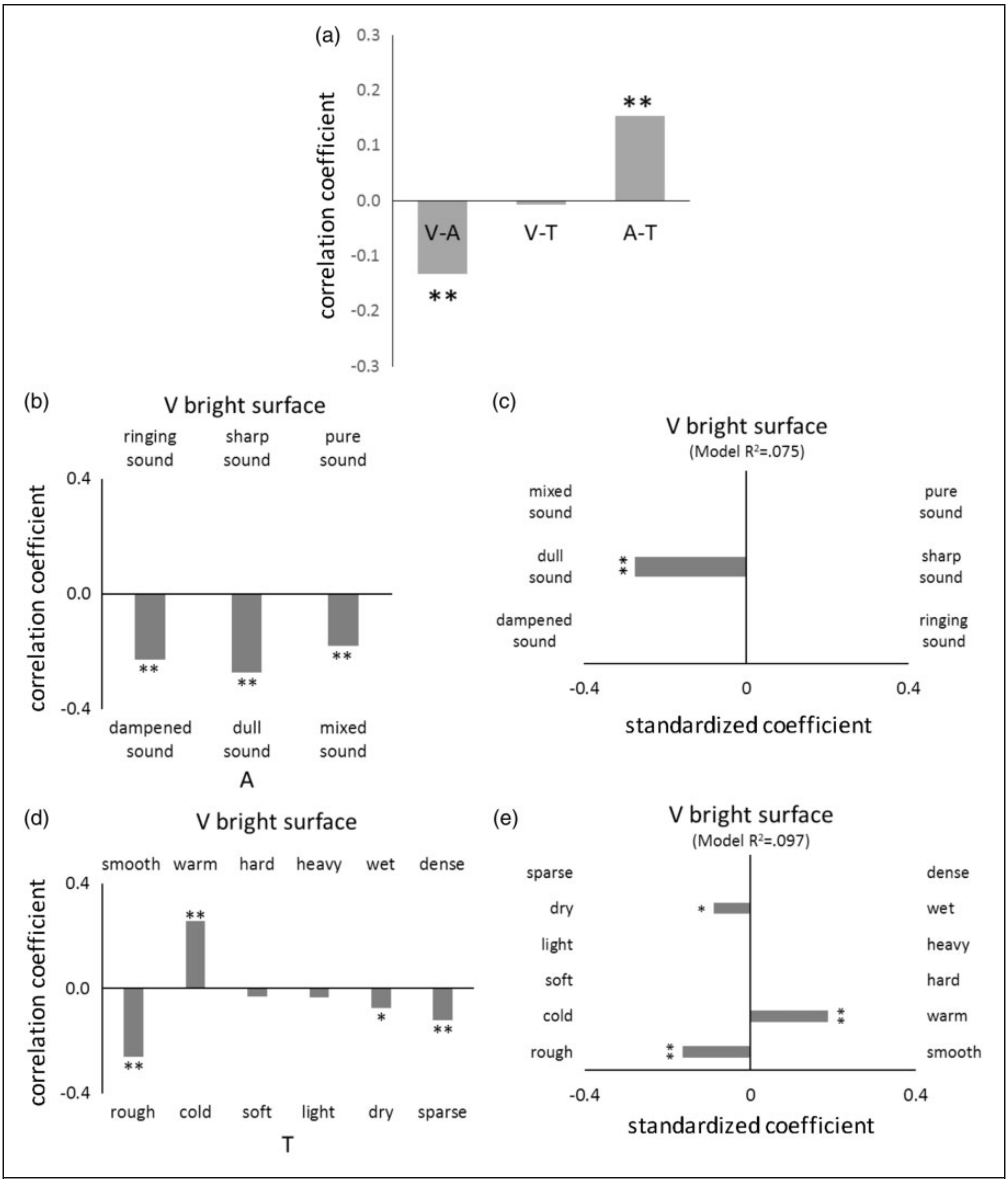

Figure 2. The results for surface brightness. (a) Correlation coefficients for ratings of "surface brightness" for the V–A, V–T, and A–T comparisons. (b) Correlation coefficients between "surface brightness" as judged by vision and three auditory properties as judged by audition. (c) The standardized coefficients obtained from multiple regression analyses with "surface brightness" as judged by vision as the dependent variable and three auditory properties as judged by audition as independent variables. (d) Correlation coefficients between "surface brightness" as judged by vision and six tactile properties as judged by touch. (e) Standardized coefficients obtained from multiple regression analyses with "surface brightness" as judged by vision as the dependent variable and six tactile properties as judged by touch as independent variables. Statistically significant values are indicated by asterisks (*, significant at $p < .05$; **, significant at $p < .01$).

by vision was the dependent variable, and "ringing of the sound", "sharpness of the sound", and "pureness of the sound" as judged by audition were the independent variables. The results are presented in Figure 2(c). The standardized relationship coefficient was significant and negative regarding "sharpness of the sound", indicating that surface brightness of wood surfaces may be predicted by dullness of emitted sound.

Further, we conducted a correlation analysis to determine the relationship of "surface brightness" as judged by vision with tactile properties such as "smoothness", "warmness", "hardness", "heaviness", "wetness", and "denseness" as judged by touch. Correlation coefficients were calculated for six comparisons between "surface brightness" and each of these tactile properties. The results are presented in Figure 2(d). A significant positive correlation was found for "warmness" ($p < .01$); significant negative correlations were found for "smoothness", "denseness" ($p < .01$ for both), and "wetness" ($p < .05$). These results suggest that wooden pieces with bright surfaces were likely to be felt as rough, warm, dry, and sparse.

Finally, we conducted a multiple regression analysis to identify tactile properties able to predict "surface brightness" as judged by vision. In this analysis, "surface brightness" as judged by vision was the dependent variable, and "smoothness", "warmness", "hardness", "heaviness", "wetness", and "denseness" as judged by touch were the independent variables. The results are presented in Figure 2(e). The standardized relationship coefficient was significant and positive for "warmness" ($p < .01$), and significant and negative for "smoothness" and "wetness" ($p < .01$ and $p < .05$, respectively). This suggests that the brightness of wood surfaces may be predicted by roughness, warmness, and dryness.

Sharpness of Sound

To determine if the degree of "sharpness of the sound" as judged by audition is consistent with "sharpness of the sound" judged by vision and touch, we conducted a correlation analysis making V–A, V–T, and A–T comparisons. The results are presented in Figure 3(a). We observed significant positive correlation in the V–T and A–T comparisons ($p < .01$ for both). The positive correlation between ratings of the sharpness of sound by audition and touch suggests that the participants correctly judged sound sharpness in the tactile condition; however, the correlation coefficient in the V–A comparison was not significant, suggesting that participants could not estimate sound sharpness accurately in the visual condition. Interestingly, a significant positive correlation was found between ratings made using vision and touch, although participants were making speculative estimations in these two modalities. This suggests that some commonalities were present between ratings made using two indirect modalities for sound sharpness.

To determine the relationship between "sharpness of the sound" as judged by audition and visual properties such as "glossiness of the surface", "brightness of the surface", and "clearness of the surface" as judged by vision, we conducted a correlation analysis. Correlation coefficients were calculated for comparisons between "sharpness of the sound" and each of the aforementioned visual properties. The results are presented in Figure 3(b). A significant negative correlation was found for "brightness of the surface" ($p < .01$); this suggests that wooden pieces with sharp sounds were likely to be judged as having dark surfaces.

We conducted a multiple regression analysis to determine which visual property predicted "sharpness of the sound" as judged by audition. In this analysis, "sharpness of the sound" as judged by audition was the dependent variable, and "glossiness of the surface", "brightness of the surface", and "clearness of the surface" as judged by vision were the independent

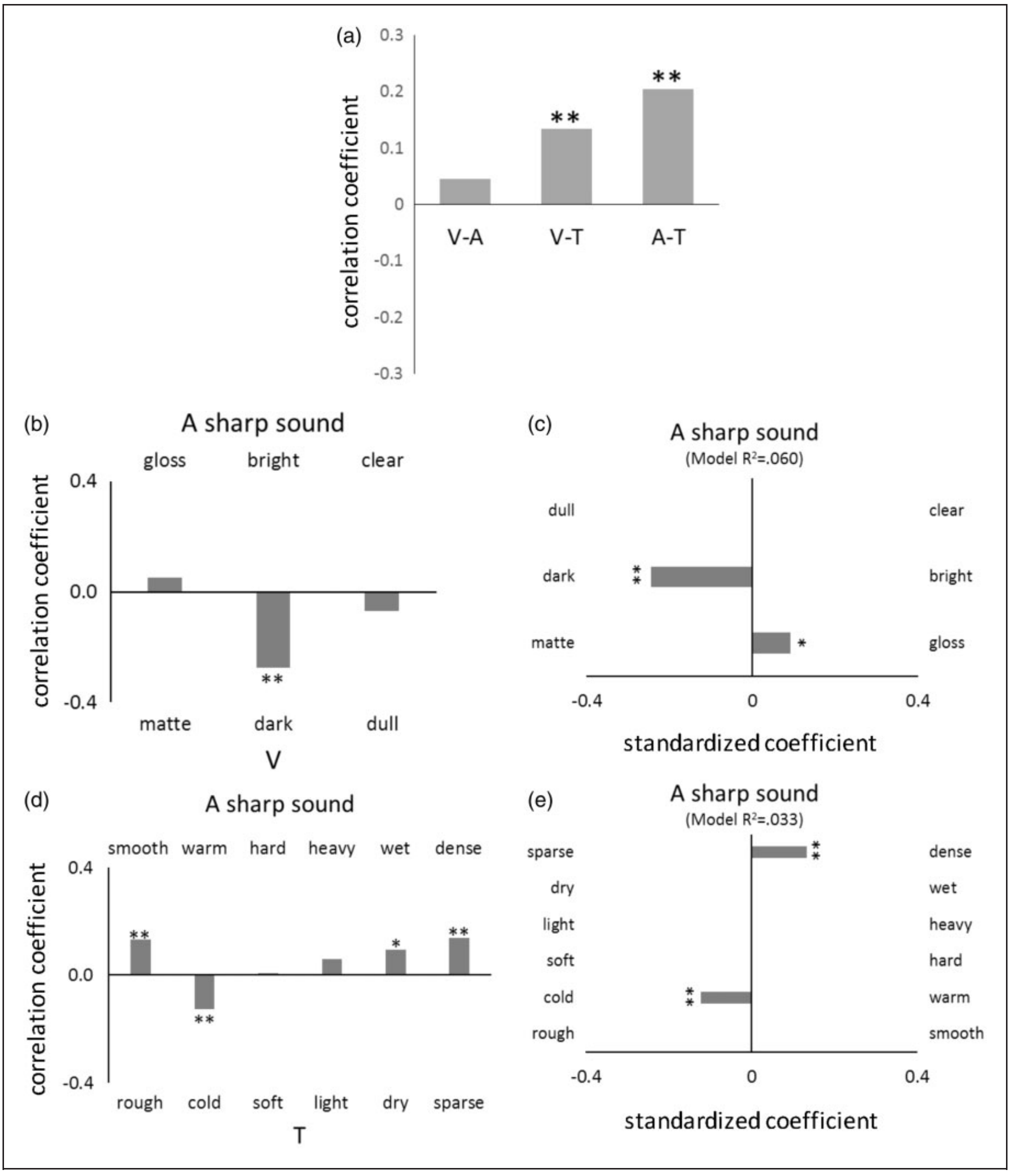

Figure 3. Results for sharpness of the sound are presented. (a) Correlation coefficients for ratings of "sound sharpness" for the V–A, V–T, and A–T comparisons. (b) Correlation coefficients between "sound sharpness" as judged by audition and three visual properties as judged by vision. (c) Standardized coefficients obtained from multiple regression analyses with "sound sharpness" as judged by audition as the dependent variable and three visual properties as judged by vision as independent variables. (d) Correlation coefficients between "sound sharpness" as judged by audition and six tactile properties as judged by touch. (e) Standardized coefficients obtained from multiple regression analyses with "sound sharpness" as judged by audition as the dependent variable and six tactile properties judged by touch as independent variables. Statistically significant values are indicated by asterisks (*, significant at $p < .05$; **, significant at $p < .01$).

variables. The results were presented in Figure 3(c). The standardized relationship coefficients were significant and positive for "glossiness of the surface" ($p < .05$), and negative for "brightness of the surface" ($p < .01$). This suggests that the sharpness of sounds made by wood may be predicted by glossiness and darkness of the surfaces.

Further, we conducted a correlation analysis to determine the relationship between "sharpness of the sound" as judged by audition and tactile properties such as "smoothness", "warmness", "hardness", "heaviness", "wetness", and "denseness" as judged by touch. The correlation coefficients were calculated for six comparisons between "sharpness of the sound" and each of the aforementioned tactile properties. The results are presented in Figure 3(d). Significant positive correlations were found for "smoothness", "denseness" ($p < .01$ for both), and "wetness" ($p < 0.5$). A significant negative correlation was found for "warmness" ($p < .01$). These results suggest that wooden pieces with sharp sounds were likely to be felt as smooth, cold, wet, and dense.

Finally, we conducted a multiple regression analysis to determine which tactile properties are able to predict "sharpness of the sound" as judged by audition. In this analysis, "sharpness of the sound" as judged by audition was the dependent variable, and "smoothness", "warmness", "hardness", "heaviness", "wetness", and "denseness" as judged by touch were the independent variables. The results are presented in Figure 3(e). The standardized relationship coefficient was significant and positive for "denseness" ($p < .01$), and significant and negative for "warmness" ($p < .01$). This suggests that the sharpness of sounds made by wood may be predicted by the denseness and coldness of its surface.

Smoothness of Surface

To determine if the degree of "smoothness" as judged by touch is consistent with "smoothness" as judged by vision and audition, we conducted a correlation analysis making V–A, V–T, and A–T comparisons. The results are shown in Figure 4(a). In this analysis, we observed a significant positive correlation in the V–T and A–T comparisons ($p < .01$ for both). The positive correlation between ratings of roughness by vision and touch suggests that participants were able to accurately predict surface smoothness in the visual condition. Moreover, the positive correlation between ratings of roughness by audition and touch suggests that the participants were also able to accurately predict smoothness in the auditory condition.

To determine the relationship between "smoothness" as judged by touch and visual properties such as "glossiness of the surface", "brightness of the surface", and "clearness of the surface" as judged by vision, we conducted a correlation analysis. Correlation coefficients were calculated for comparisons between "smoothness" and each of these visual properties. The results are presented in Figure 4(b). The correlation coefficients were significant and positive for "glossiness of the surface" ($p < .05$), and significant and negative for "brightness of the surface" ($p < .01$). This suggests that wooden pieces that were smooth to the touch were likely to be judged as having glossy and dark surfaces.

We conducted a multiple regression analysis to identify visual properties able to predict "smoothness" as judged by touch. In this analysis, "smoothness" as judged by touch was the dependent variable, and "glossiness of the surface", "brightness of the surface", and "clearness of the surface" as judged by vision were the independent variables. The results are presented in Figure 4(c). The standardized relationship coefficients were significant and positive for "glossiness of the surface" and "clearness of the surface" ($p < .01$ for both), and significant and negative for "brightness of the surface" ($p < .01$). This means that the smoothness of a wood surface may be predicted by its glossiness, clearness, and darkness.

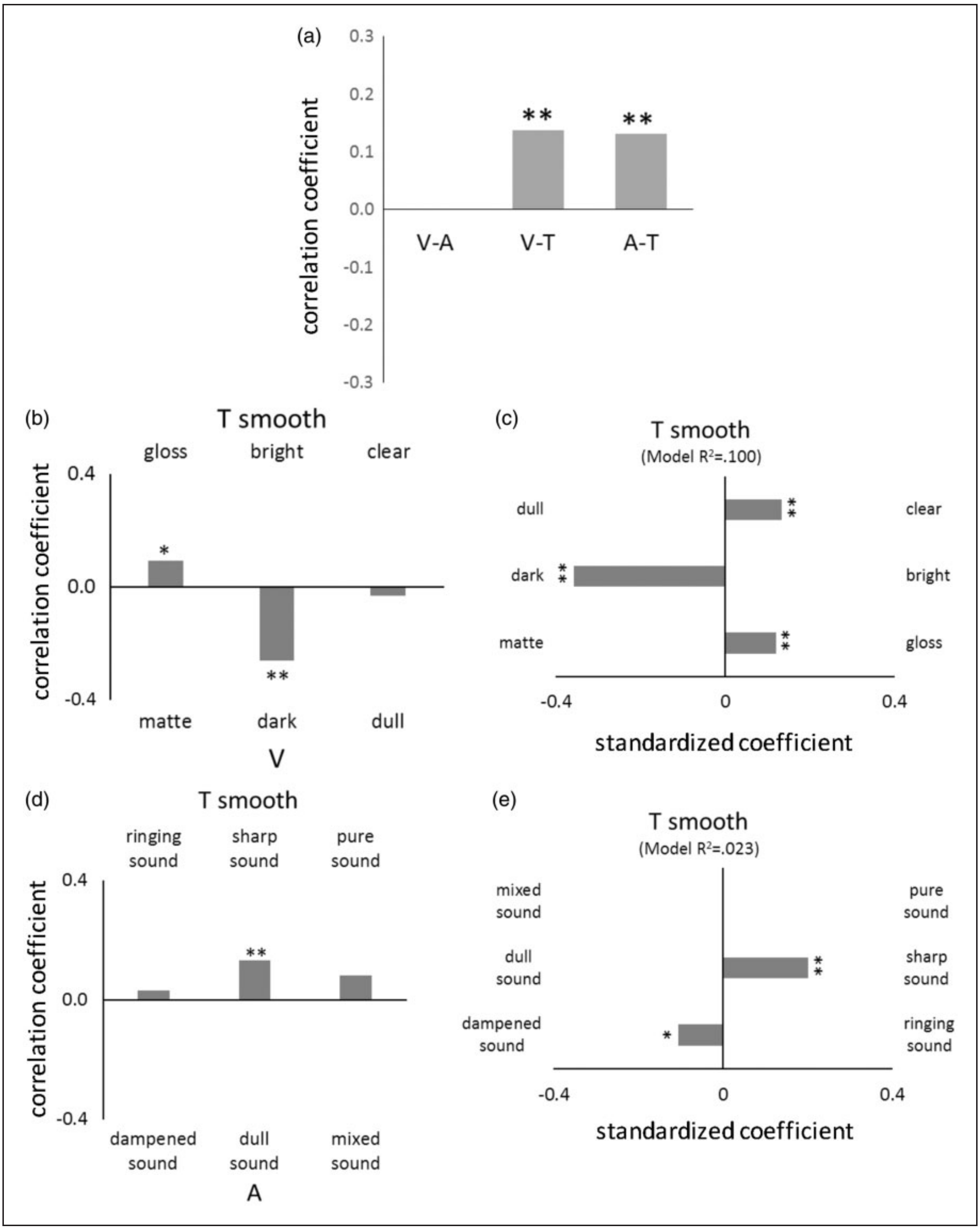

Figure 4. Results for smoothness are presented. (a) Correlation coefficients for ratings of "smoothness" for V–A, V–T, and A–T comparisons. (b) Correlation coefficients between "smoothness" as judged by touch and three visual properties as judged by vision. (c) Standardized coefficients obtained from multiple regression analyses with "smoothness" as judged by touch as the dependent variable and three visual properties as judged by vision as independent variables. (d) Correlation coefficients between "smoothness" as judged by touch and three auditory properties as judged by audition. (e) Standardized coefficients obtained from multiple regression analyses with "smoothness" as judged by touch as the dependent variable and three auditory properties as judged by audition as independent variables. Statistically significant values are indicated by asterisks (*, significant at $p < .05$; **, significant at $p < .01$).

Further, we conducted a correlation analysis to determine the relationship between ''smoothness'' as judged by touch and auditory properties such as ''ringing of the sound'', ''sharpness of the sound'', and ''pureness of the sound'' as judged by audition. Correlation coefficients were calculated for comparisons between ''smoothness'' and each of these auditory properties. The results are presented in Figure 4(d). The correlation coefficient was significant and positive for ''sharpness of the sound'' ($p < .01$). This suggests that wooden pieces that were smooth to the touch were likely to be judged as making sharp sounds.

Finally, we conducted a multiple regression analysis to identify auditory properties able to predict ''smoothness'' as judged by touch. In this analysis, ''smoothness'' as judged by touch was the dependent variable, and ''ringing of the sound'', ''sharpness of the sound'', and ''pureness of the sound'' as judged by audition were the independent variables. The results are presented in Figure 4(e). The standardized relationship coefficient was significant and positive for ''sharpness of the sound'' ($p < .01$), and significant and negative for ''ringing of the sound'' ($p < .05$). This suggests that the smoothness of a wood surface may be predicted by sharp and dampened sounds.

Summary of the Results

The results of the first correlation analysis (a) in Figures 2–4 are summarized as follows. Significant positive correlations were found between ratings made by audition and touch across all three properties. Significant positive correlations were also found between ratings made by vision and touch for two of the three properties (sharpness of the sound and smoothness). Nonetheless, no positive correlation was found between ratings made by vision and audition for any properties.

Results of the second correlation analysis (b and d) and the multiple regression analysis (c and e) in Figures 2–4 exhibited a similar tendency, although small differences were visible. The findings are as follows. Wood pieces that were judged as bright by vision were likely to be judged as making dampened, dull, and mixed sounds by audition, and as feeling rough, warm, dry, and sparse by touch. Wood pieces judged as making a sharp sound were likely to be judged as glossy and dark by vision, and smooth, cold, wet, and dense by touch. Wood pieces that were smooth to the touch were likely to be judged as glossy, dark, and clear by vision, and as making dampened and sharp sounds by audition.

Discussion

The present study reanalyzed data collected by Fujisaki et al. (2015), examining perceptions of the material properties of wood using vision, audition, and touch. We compared participant evaluations of three perceptual properties (surface brightness, sharpness of sound, and smoothness) of the wood blocks obtained separately via vision, audition, and touch. If the human perceptual system learns co-occurrences of all multi-sensory signals in the natural world, participant evaluations would be consistent for every combination of modalities; however, the present study's results indicate that evaluations are only consistent for some combinations. It may, therefore, be the case that the statistical learning of co-occurrences is only possible under certain conditions.

First, the surface features of an object are primarily perceived by vision, and the properties of material inside an object is detected by audition, while both these pieces of information may be detected by touch[3] (Lederman & Klatzky, 2009). Therefore, vision and touch may receive surface information in common, and audition and touch may receive information about the material inside objects in common. It is, therefore, possible that the statistical learning of co-occurrences is efficient only for the V–T and A–T combinations. By contrast, learning may be inefficient for the V–A combination, as these modalities receive little information in common. Additionally, vision and audition typically receive sensory input passively from distal stimuli, while touch often actively obtains sensory input from proximal stimuli. It is, therefore, possible that we particularly accomplish statistical learning for the V–T and A–T combinations, as touch's active exploration of the external world facilitates learning. Participant evaluations were particularly consistent for the A–T combination. This may reflect the fact that the peripheral mechanisms of audition and touch are relatively similar in that they both arise from stimulation of peripheral mechanoreceptors; further, these two modalities' compatibility has been demonstrated by various types of psychophysical tasks (Butler, Foxe, Fiebelkorn, Mercier, & Molholm, 2012; Fujisaki & Nishida, 2009; Soto-Faraco & Deco, 2009; Yau, Olenczak, Dammann, & Bensmaia, 2009).

The results of the correlation and multiple regression analyses (b–e in Figures 2–4) accorded with externally existing trends in the natural world. Regarding wood, samples with low specific gravity generally have low density. These are likely to be judged as, for example, warm and making dull sounds, due to physical properties such as thermal conductivity and their manner of transmission of sound waves. Although surface brightness is not directly determined by specific gravity (Moya, Fallas, Bonilla, & Tenorio, 2012), many species of wood with low specific gravity have relatively bright surfaces in the natural world. This also obtained for the sample blocks used in Fujisaki et al. (2015). The results presented in Figures 2–4 accord with these trends. This suggests that we learn the objective co-occurrence of multi-sensory signals in the natural world and are subsequently able to make speculative judgments about certain key material properties of objects using information from indirect modalities.

Cross-modal correspondence is most commonly reported between vision and audition. This may be partly accounted for by the results of the correlation and multiple regression analyses (b–e), as presented in Figures 2–4; however, this is incompatible with the lack of significant positive correlation in the V–A comparison in the correlation analysis (a) in Figures 2–4. This may be due to the characteristics of the data used in the analyses in this study. Most cross-modal correspondence between vision and audition is related to the sizes and shapes of visual stimuli (Evans & Treisman, 2010; Parise & Spence, 2012), properties that are most easily detected by vision; however, Fujisaki et al. (2015) did not ask participants about the sizes and shapes of their visual stimuli, and instead used stimuli of identical sizes and shapes. If the stimuli had been of varying sizes and shapes, higher positive correlation coefficients may have been observed in the V–A comparisons.

Further, the values of the correlation and standardized coefficients were generally small. This may also be due to the characteristics of the stimuli. The stimuli used were not optimized for the aim of this study, as we reanalyzed data obtained by Fujisaki et al. (2015). Their stimuli were wood blocks that were controlled for sizes and shapes, and for surface texture, as they were sanded with the same grit sandpaper. Many perceptual properties were, therefore, relatively uniform among the stimuli, potentially making co-variations between the various modalities difficult to detect. If stimuli that varied in materials and features were used, co-variations between sensory modalities may become easier to detect. Future research should, therefore, replicate the present study's aims in examination of natural objects made various materials. Further, it would be interesting to examine correlations of participants' evaluations across all adjective pairs used. In the first correlation analysis in the present study (Figures 2a, 3a, and 4a), we focused on three perceptual properties because they exhibited the highest inter-participant correlation coefficients (IPC) when rated with their 'direct' modalities in Fujisaki et al. (2015). Although we did so in order to simplify the results to determine essential trends in the data, a more accurate description of our cross-modal material perception can be obtained from correlations between ratings for a range of properties. This analysis should be performed in a future study that uses stimuli with sufficient variation in multiple properties and that show relatively constant IPC values over multiple adjective pairs.

As this study's results support the assumption that we learn correlations between multi-sensory signals from experience in the natural environment, it is important to address how such learning arises through experience. For instance, two extreme alternative hypotheses are as follows. If multi-sensory signals derived from natural objects are always correlated in some fixed way, cross-modal correlations in a given object may be predictable without any previous experience of that object. By contrast, multi-sensory experience of an object is necessary for estimating cross-modal correlations in that object. The present study cannot directly test these hypotheses, as Fujisaki et al. (2015) did not aim to control participants' experiences of the stimuli. Nonetheless, our finding that stronger cross-modal correspondence was observed in the A–T and V–T comparisons suggests that feedback resulting from active exploration of the external world facilitates statistical learning. Moreover, we currently assume that these two hypotheses are not mutually exclusive, and that we indeed learn objectively present co-variations through experience. Future research should address this question directly, for example, by using unfamiliar stimuli that incorporate co-variations in themselves.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by research funds from Sumitomo Forestry Co. Ltd. WF is partially supported by JSPS KAKENHI Grant Numbers 22135004, 2538099 and 16H01685. SK is partially supported by Grant-in Aid for JSPS fellows(15J01103).

Notes

1. The reason why there were two sets of stimuli (Series 1 and Series 2) was that we aimed to examine the properties for each tree species, not the properties specific to each sample block.
2. The species of tree were as follows: Japanese cedar, Japanese cypress, pine, falcata, poplar, lauan, maple, chestnut, walnut, cherry, oak, teak, guibourtia, and ebony.
3. Surface textures, such as roughness, and characteristics of the material inside, such as stiffness and thermal quality, are both perceived by touch.

References

Asano, M., & Yokosawa, K. (2013). Grapheme learning and grapheme-color synesthesia: Toward a comprehensive model of grapheme-color association. *frontiers in Human Neuroscience*, *7*, 757.

Ben-Artzi, E., & Marks, L. E. (1995). Visual–auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*, 1151–1162.

Bennett, C. A., & Rey, P. (1972). What's so hot about red? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *14*, 149–154.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*, 241–247.

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*, 663–672.

Burr, D. C., Parise, C. V., & Spence, C. (2009). 'When Birds of a Feather Flock Together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, *4*, e5664.

Butler, J. S., Foxe, J. J., Fiebelkorn, I. C., Mercier, M. R., & Molholm, S. (2012). Multisensory representation of frequency across audition and touch: High density electrical mapping reveals early sensory-perceptual coupling. *The Journal of Neuroscience*, *32*, 15338–15344.

Crisinel, A.-S., Jacquier, C., Deroy, O., & Spence, C. (2013). Composing with cross-modal correspondences: Music and odors in concert. *Chemosensory Perception*, *6*, 45–52.

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

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*, 878–896.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 6.

Fujisaki, W., & Nishida, S. y. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. *Experimental Brain Research*, *198*, 245–259.

Fujisaki, W., Tokita, M., & Kariya, K. (2015). Perception of the material properties of wood based on vision, audition, and touch. *Vision Research*, *109*, 185–200.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*, 1191–1203.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*, 274–284.

Ho, H.-N., Van Doorn, G. H., Kawabe, T., Watanabe, J., & Spence, C. (2014). Colour-temperature correspondences: When reactions to thermal stimuli are influenced by colour. *PLoS One*, *9*, e91854.

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

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, *9*, 399–410.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*, 1439–1459.

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

Marks, L. E., Hammeal, R. J., Bornstein, M. H., & Smith, L. B. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, *52*, 1–102.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*, 903–923.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 69–79.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*, 323–336.

Mogensen, M. F., & English, H. B. (1926). The apparent warmth of colors. *The American Journal of Psychology*, *37*, 427–428.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 133–136.

Moya, R., Fallas, R. S., Bonilla, P. J., & Tenorio, C. (2012). Relationship between wood color parameters measured by the CIELab system and extractive and phenol content in Acacia mangium and Vochysia guatemalensis from fast-growth plantations. *Molecules*, *17*, 3639–3652.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *442*, 257–261.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*, 319–333.

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*, 755–775.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of consciousness studies*, *8*, 3–34.

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

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

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*, 99–142.

Soto-Faraco, S., & Deco, G. (2009). Multisensory contributions to the perception of vibrotactile events. *Behavioural Brain Research*, *196*, 145–154.

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

Velasco, C., Jones, R., King, S., & Spence, C. (2013). The sound of temperature: What information do pouring sounds convey concerning the temperature of a beverage. *Journal of Sensory Studies*, *28*, 335–345.

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal frequency channels are linked across audition and touch. *Current Biology*, *19*, 561–566.