

Irrelevant visual faces influence haptic identification of facial expressions of emotion

Roberta L. Klatzky · Aneta Abramowicz ·
Cheryl Hamilton · Susan J. Lederman

Published online: 9 November 2010
© Psychonomic Society, Inc. 2010

Abstract This study demonstrates that when people attempt to identify a facial expression of emotion (FEE) by haptically exploring a 3D facemask, they are affected by viewing a simultaneous, task-irrelevant visual FEE portrayed by another person. In comparison to a control condition, where visual noise was presented, the visual FEE facilitated haptic identification when congruent (visual and haptic FEEs same category). When the visual and haptic FEEs were incongruent, haptic identification was impaired, and error responses shifted toward the visually depicted emotion. In contrast, visual emotion labels that matched or mismatched the haptic FEE category produced no such effects. The findings indicate that vision and touch interact in FEE recognition at a level where featural invariants of the emotional category (cf. precise facial geometry or general concepts) are processed, even when the visual and haptic FEEs are not attributable to a common source. Processing mechanisms behind these effects are considered.

Keywords Facial expression of emotion · Emotion perception · Face perception · Haptic · Multi-sensory · Irrelevant vision

Introduction

A basic component of social communication is the modulation of facial expression in response to emotion

R. L. Klatzky (✉)
Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA 15213-3890, USA
e-mail: klatzky@cmu.edu

A. Abramowicz · C. Hamilton · S. J. Lederman
Department of Psychology, Queen's University,
Kingston, ON, Canada

(Darwin, 1872/1955). The visible cues from the face have been extensively studied, and the foundations of six universally recognized emotional expressions in the facial musculature have been documented in detail (Ekman et al., 1987; Ekman & Friesen, 1975).

The signaling of emotion by modalities other than vision has also been of considerable interest, particularly the role of auditory cues (especially speech). Facially and vocally communicated emotions appear to be integrated automatically and early in processing (see de Gelder, Vrommen, & Pourtois, 1999, for review). Massaro and Egan (1996) found that the relative influence of the two modalities followed the fuzzy-logical model of perception, such that the extent to which one was used depended on the quality of the other. Other research suggests that the relative reliance on voice versus face is age-related, with the relative contribution of vision increasing during infancy (Heyman, 1996), and again from early childhood to adulthood (Zupan, 2008).

What is less well recognized is that the cues that arise from facial expressions are also available to the sense of touch, and hence, facial expressions of emotion (FEEs) can be recognized by the haptic perceptual system. Lederman, Klatzky, Abramowicz, and associates (2007) showed that the six basic FEEs could be classified at levels well above chance when people felt the static expression or its dynamic formation and cessation by a live actor (classification accuracies of 51 and 74%, respectively, cf. 17% chance). High accuracy (81%) was also found for haptic classification of FEEs in rigid facemasks molded from live actresses (Baron, 2008).

Not only do the visual and haptic perceptual systems both show capability for recognizing facial expression, but there is also evidence for similarity in the underlying mechanisms (for review, see Lederman, Klatzky, & Kitada,

2010). The well-known phenomenon that inversion undermines visual face recognition (e.g., Diamond & Carey, 1986; Yin, 1969) is paralleled by inversion effects for recognition of FEEs in faces, and this is true for both vision (Direnfeld, 2007, live faces) and touch (Lederman et al., 2008; 2D raised-line drawings; but cf. a null inversion effect for haptically perceived facemasks in Baron, 2008). Moreover, neuroimaging studies of responses to FEEs have found commonalities across the two modalities. Kitada, Johnsrude, Kochiyama, and Lederman (2010) directly compared brain activation induced by visual and haptic FEEs relative to control objects. They found activation unique to emotional expressions by both modalities in areas believed to process information about actions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Montgomery & Haxby, 2008), including the inferior frontal gyrus, inferior parietal lobe, and regions of the superior temporal sulcus.

The findings that vision and touch invoke similar mechanisms and cortical structures for processing FEEs suggest a strong potential for interaction between the modalities. Such interactions have been of interest since early demonstrations that simultaneous vision could essentially override inputs from touch about the same source event (e.g., Rock & Victor, 1964). Subsequent models of visual/haptic interactions have assumed that the visual channel is given a greater weight than the haptic, by virtue of its higher reliability (Ernst & Banks, 2002) or general "modality appropriateness" (Welch & Warren, 1980).

The issue of cross-modal involvement in face processing has not been raised heretofore with respect to classification of emotion, although vision/touch interactions have been investigated in regard to facial identity. That individual faces can be recognized by touch, and even show inversion effects similar to vision, was first demonstrated by Kilgour and Lederman (2006). These authors also assessed performance in a cross-modal transfer task, in which faces studied with one modality were tested in another, and found little evidence of transfer. Casey and Newell (2007) also used the transfer task and found that, while cross-modal identification accuracy was above chance, there was a cost relative to unimodal study and test, regardless of the direction of transfer across modalities. Moreover, data were also reported from a preliminary study in which faces were learned bimodally and then tested unimodally or bimodally. Unimodal haptic performance was relatively poor, and there was no advantage for bimodal over unimodal visual recognition. The authors concluded that, when visual information is present, it dominates facial processing. Similar conclusions were reached by Dopjans, Wallraven, and Bülthoff (2009), who found that when both study and test were in the visual modality, recognition performance was clearly superior to unimodal haptic and cross-modal

visual/haptic conditions, which were all equivalent. In particular, the haptic recognition process seemed unable to capitalize on the greater bandwidth arising from using the visual system for the study phase.

Other paradigms further support a decoupling, rather than integration, between visual and haptic face recognition. Congenital and early-blind individuals have been found to achieve high accuracy in successive matching of face masks (Pietrini et al., 2004), indicating that visual experience, and the mediation it offers, is not necessary for haptic face processing. This conclusion is reinforced by Kilgour and Lederman (2002) finding a low correlation between the rated vividness of visual images and the rate of identification of 3D facemasks by touch. Also relevant is an imaging study by Kitada, Johnsrude, Kochiyama, and Lederman (2009), which concluded that, although at a coarse anatomical level, the fusiform gyrus shows sensitivity to both haptic and visual facial stimuli, the modalities differ with respect to functional architectures for recognition of facial identity. Functional and neural differences between visual and haptic face processing may reflect differential reliance on what Buck (1984) called direct and mediated decoding. Buck's particular interest was in recognizing facial emotion, in which case direct decoding is presumably linked to the same fundamental processes that induce spontaneous emotional expressions, whereas mediated decoding relies on learned rules about the associated facial patterns.

In short, the literature on visual/haptic interactions in face identification is somewhat equivocal on the extent to which the modalities interact, but offers relatively little support for integrative processes. This raises the question of whether, and if so how, inputs from simultaneously seen and touched faces might interact, and more specifically, whether such interactions necessitate attribution to a common source. It should be noted that, most typically in studies of cross-modal interaction during stimulus encoding, the expectation is induced that all modalities convey information about the same stimulus, and physical discrepancies are meant to be minimally detectable. Failure to attribute the bimodal information to the same physical event undermines its joint use (Helbig & Ernst, 2007). However, there is also evidence that clearly irrelevant visual input can influence haptic processing. For example, sight of a stimulated body part improves spatial acuity (Kennett, Taylor-Clarke, & Haggard, 2001), noninformative vision enhances haptic judgments of parallelity (Newport, Rabb, & Jackson, 2002), and an irrelevant flash of light at the location of a vibro-tactile stimulus influences threshold discrimination of vibration (Arabzadeh, Clifford, & Harris, 2008).

In the present paper, we tested whether simultaneous emotional cues from a visual face would influence the classification of emotion from a haptically explored face-mask, even when the haptic and visual faces were clearly

discrepant. If so, this would extend the context of influence for irrelevant vision and would suggest a novel basis for visual/haptic interactions in face processing. In our task, participants were instructed to identify a haptically depicted expression while merely looking at (but not classifying) a pictured face displaying a FEE. There were clear cues that the visual and haptic FEEs stemmed from different sources: the facemasks and pictures were not co-located, and they portrayed different people. More often than not, the two stimuli also depicted different FEE categories. The question was whether the visual display, which carried emotional information but was irrelevant to the assigned task, would nevertheless influence processing of the haptically explored face.

At the outset, we considered that influence of an irrelevant visual FEE on haptic FEE classification could arise from different levels of correspondence between the stimuli. First, visual and haptic faces correspond to the extent that their physical geometry is congruent. When FEEs are portrayed visually and haptically by the same person making the same expression, matching can occur at this *physical* level. However, as the literature on FEEs emphasizes, emotional faces match to the extent that they portray general feature patterns characteristic of the same emotion, such as the upturned lips that characterize happiness (Ekman & Friesen, 1975). At this *categorical* level, faces portraying the same FEE need not be structurally congruent. Rather, the commonalities arise because the features are shaped by the underlying facial musculature as the emotion is felt, and this happens in the same way across individuals. Finally, to the extent that FEEs correspond to meaningful categories with labels, visual and haptic interactions could be mediated by commonalities at an abstract *conceptual* level.

Of these three levels, our interest is in the categorical level, where cross-modal interactions presumably reflect processing of facial emotion signals that are shared across individuals, rather than precise geometry or common concept. To ensure that any effects of visual/haptic face congruence in the present task were not due to matching or mismatching of the emotion at the physical level, the visual and haptic FEEs were portrayed by different actors. As a further control for the effect of matching or mismatching at the level of the emotion concept, as opposed to emotional content in the particular face displays, we compared the effect of visually displaying a FEE during haptic classification to the effect of displaying an emotion label.

Three outcomes were assessed: a visual FEE could facilitate haptic processing of a face signaling the same emotional category (i.e., when congruent), it could impair haptic processing when the two emotions mismatched (i.e., when incongruent), and/or it could bias decisions, thus shifting responses toward the visually displayed category.

To measure these effects, the simultaneous-face conditions were compared to a control condition in which visual noise was presented during haptic FEE classification. More specifically, the task produced three empirical measures, corresponding to the three outcomes just described. The *congruence effect* is defined as the accuracy (proportion correct) on congruent trials (visual and haptic FEEs same) minus the accuracy on control trials (visual noise), whereas the *incongruence effect* is control accuracy minus accuracy on incongruent trials (visual and haptic FEEs differ). The third empirical measure came from the distribution of errors on trials where an incongruent visual FEE is presented; specifically, it measures the tendency of errors to match the visual FEE more than would be expected by chance. Because this *visual-match tendency* is expressed as a proportion of the total number of errors, the measure is independent of the magnitude of the incongruence effect.

To the extent that these effects occur at the conceptual level, they should be at least as great when labels (which are unambiguous) are viewed as when faces are viewed. For example, the visual-match tendency might simply reflect a bias to respond with the name of the visually displayed emotional category. If so, displaying the name by itself should be a powerful stimulus. To the contrary, a larger visual-match tendency for faces than FEE labels would indicate that the actual presence of FEE features, not just the conceptual identity of the visual FEE, is needed in order for vision to influence haptically based responses. By including the label control and using different individuals for the haptic and visual FEEs, the present experiments sought evidence of cross-modal interactions at the level of features that are signals of emotion across individuals. Moreover, given that the FEE identification was to be based on the haptic stimulus, while the visual stimulus was merely present and clearly discrepant, such effects would suggest that vision/touch interactions were not driven by the attribution to a common bimodal source.

Method

Part A: Faces

The experimental task was to identify the FEE portrayed by a haptically presented 3D facemask while viewing a photograph of either the same FEE (congruent), a different FEE (incongruent), or a neutral control stimulus (random-dot noise). The participant was instructed to respond with the name of the FEE portrayed by the haptic stimulus, regardless of the emotion depicted visually, and to maintain gaze on the monitor while exploring the mask. No indication was given of the potential relation between visual and haptic displays, and no feedback was provided.

Subjects Twenty-five participants from the university population took part. There were 8 males and 17 females (average age 18.6 years), relative proportions that are representative of the population sampled. All gave informed consent.

Stimuli The haptic stimuli were 12 3D facemasks, 2 each displaying the facial expressions of anger, disgust, fear, happiness, sadness, and surprise. A practice trial used a mask with a neutral expression. To create the facemasks, two trained female actors (29 and 69 years) depicted the FEEs with their eyes closed as a Cyberware Color 3D Digitizer Model 3030R6B/PS scanned their faces. Anomalies detected by the scanner were airbrushed from the images, and the 3D facemasks were created in white ABS plastic using the Dimension SST 3D printer. Representative examples are shown in Fig. 1. Note that, although the female actors differed in age, previous studies using masks of their faces have found no age effects on identification (Lederman et al., 2007; raised-line drawings: Lederman et al., 2008; 3D facemasks: Baron, 2008).

The visual stimuli were grayscale photographs of the same 6 FEEs (from Ekman, Friesen, & Hager, 2002). Each FEE was portrayed by two actresses. The photographs were rendered 19.0 cm high and 12.5 cm wide on a 43.2-cm screen with a black background. A rectangle of the same size as the photographs, portraying random-dot grayscale visual noise, was used for control trials.

A preliminary experiment assessed the visual identifiability of the FEEs from the photographs. Five participants, drawn from the same pool as the experiment proper but who did not take part in that study, were visually shown and identified the FEEs in the 12 photographs over six blocks of trials, with each photograph presented once in random order within each block. Averaging across actresses, the percentage correct visual identification was:

anger 84%, disgust 82%, fear 52%, happiness 100%, sadness 90%, and surprise 92% (average 83%). Although the expressions differed significantly in recognition rates, $F(4, 16) = 7.16$, $p = .002$, all were identified significantly above chance. For comparison, untrained observers were found to visually identify FEEs from facemasks with accuracy averaging 72% (Klatzky, Dierenfeld, Baron, Hamilton, & Lederman, 2010).

Procedure and design The experimenter verbally repeated the six possible emotion responses, in alphabetical order, at the beginning of the experiment and every 20 trials thereafter, as well as at the participant's request.

The participant wore LCD goggles with shutter lenses and rested his or her chin on a support at a viewing distance of 55 cm from the screen. At the start of each trial, the facemask was placed on a table in front of the participant, who made a triangle shape with the fingers of the two hands, by touching both thumbs together, and likewise both forefingers. The experimenter guided the hands into a position directly above the mask and gave a "go" signal, at which the participant dropped the hands to explore the mask. At contact, the experimenter cleared the LCD glasses, exposing the face picture, and started a timer. When the participant verbally responded with the FEE name, he or she simultaneously raised both hands, at which point the glasses were rendered opaque and the timer stopped. Participants were told to respond as quickly but as accurately as possible.

The experiment consisted of 60 trials in random order, corresponding to 3 congruent trials (same visual and haptic FEE), 4 incongruent trials (different visual and haptic FEE), and 3 control trials (visual noise stimulus) for each of 6 haptic FEEs. The proportion of congruent trials was held to 30% so as not to assume matching or encourage guessing



Fig. 1 Examples of facemasks depicting (left to right) happiness, anger, and sadness, respectively

with the name of the visual FEE. There was an initial practice trial, without vision, during which the participant freely explored the neutral-FEE mask to familiarize him or her with the general nature of the 3D facemasks.

The following practices were adopted to counterbalance use of the stimuli: On congruent trials, given that there were two actresses per visual and haptic FEE, there were four possible actress combinations for each FEE. Within each participant, across the three congruent trials per FEE, three different combinations out of the four were randomly chosen. Across participants, all four combinations were used as equally as possible given the n of 25. For the control condition using visual noise, there were two possible actresses for each haptic FEE. Across the three control trials per FEE, one actress was used twice within a participant and one was used once. Across participants, both actresses were used as equally as possible for each FEE. On the four incongruent trials for each haptic FEE category, there were 20 possible combinations of haptic and visual stimuli (2 haptic actresses \times 2 visual actresses \times 5 incongruent FEEs). From these, four combinations were randomly chosen for each participant.

Part B: Words

Subjects Twenty-five participants from the same university population took part; none had participated in Experiment 1A. There were 5 males and 20 females (average age 18.4 years). All gave informed consent.

Stimuli The haptic stimuli and the visual control (random-dot pattern) were the same as those used in Experiment 1A. The visual stimuli were light-grey labels of the six FEEs in 60 point Arial type on a black background. The height of the letters was approximately 2.5 cm, and the width varied from 1.3 to 3.2 cm, so that the longest word (surprise) was approximately 10.2 cm wide.

Procedure and design These were identical to Experiment 1A.

Results

Accuracy for congruence and incongruence effects

Figure 2 shows the mean (and SEM) proportion correct by FEE and visual display (congruent, incongruent, noise), for Experiment 1A (visual faces) and Experiment 1B (words). Relative to the accuracy of .53 for noise stimuli in Experiment 1A, congruent visual faces enhanced propor-

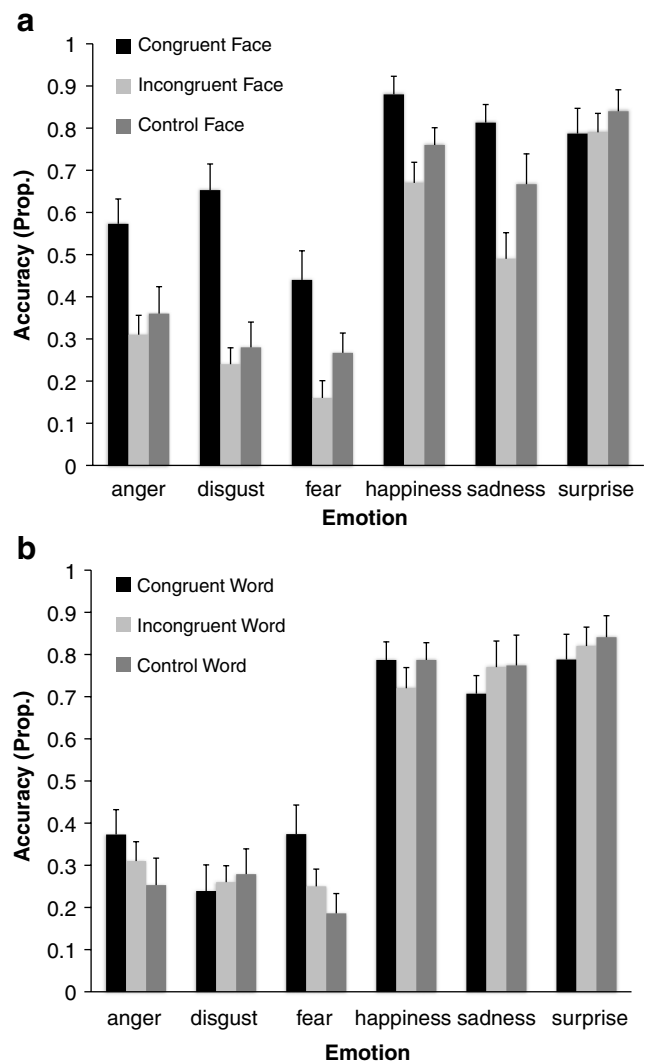


Fig. 2 Accuracy by display type (congruent, incongruent, control) and FEE for face stimuli (Experiment 1a: top) and word stimuli (Experiment 1b: bottom). Error bars +1SEM. Corresponding data for the visual faces depicting the same FEEs are presented in the text

tion correct by .16, or 30% of the control level (congruence effect), and incongruent visual faces impaired performance by .09, or 17% of control (incongruence effect). Corresponding effects for words in Experiment 1B were not evident: accuracy for noise stimuli in that study averaged .52, accuracy with congruent FEE labels was only .02 (4% of control) greater, and no impairment due to incongruent labels was observed.

An initial ANOVA was conducted, in which Visual Stimulus (face, label) was a between-subject factor along with the within-subject factors of Display Type (3 levels: congruent, incongruent, control) and FEE (6 levels). The interaction between Stimulus and Display was significant, $F(2, 96) = 14.88, p < .001, \text{partial } \eta^2 = .24$. Pairwise comparisons between levels of Visual Stimulus, using the LSD method,

$p < .05$, showed that using faces as the visual stimulus produced significantly greater accuracy for the congruent condition and significantly lower accuracy for the incongruent condition. There was no effect of the Visual Stimulus factor on the control condition; that is, haptic FEE identification when viewing visual noise was not affected by whether visual faces or labels were presented on other trials. Effects of FEE will be discussed in the context of ANOVAs described next.

Given the interaction with Visual Stimulus, further ANOVAs were conducted separately on the face and label data from Experiments 1A and 1B, respectively. The factors for each ANOVA were Display Type and FEE, both within-subject. For Experiment 1A, the ANOVA found significant effects of Display, $F(2, 48) = 30.38, p < .0001$, partial $\eta^2 = .56$, and of FEE, $F(5, 120) = 32.22, p < .0001$, partial $\eta^2 = .57$. The interaction was also significant, though small in effect size, $F(10, 240) = 3.68, p < .0001$, partial $\eta^2 = .13$. Pairwise comparisons with the LSD method, $p < .05$, 2-tailed, found that all three display conditions differed significantly overall. When Display was considered by FEE, LSD tests showed that the incongruent condition was significantly worse than the control for fear, happiness, and sadness; the congruent was significantly better than the control condition for anger, disgust, and fear; and the congruent condition significantly exceeded the incongruent condition for all FEEs but surprise (which produced the highest performance). As has been found in previous studies, the FEEs for anger, disgust, and fear all showed significantly lower accuracy than happiness, sadness, and surprise. In addition, there were significant differences between disgust and fear, and between sadness and surprise.

For Experiment 1B using labels, the corresponding ANOVA found no effect of Display, $F(2, 48) < 1$, but the effect of FEE was significant, $F(5, 120) = 44.49, p < .0001$, partial $\eta^2 = .65$. The interaction was not significant, $F(10, 240) = 1.61, p = .10$, partial $\eta^2 = .06$. Pairwise comparisons with the LSD method, $p < .05$, found that the FEEs for anger, disgust, and fear all differed from happiness, sadness, and surprise; no other pairs differed significantly.

We further examined the accuracy data for the pattern of confusions between FEEs. Six confusion matrices were generated, one for each combination of Display Type and Visual Stimulus. Each cell in a matrix corresponded to the proportion of total responses naming a given FEE category that occurred to a given stimulus FEE (this normalization corrects for differential response rates across FEEs). Although incongruence effects could moderate the confusion pattern, shifting errors toward the visual face (as confirmed in the analysis below), correlations computed across the off-diagonal cells (i.e., the confusion errors) between each pair of matrices were generally significant (over the 15 correlations, the $r(28)$ ranged from .74 to .94 and averaged .86, all $ps < .001$).

Response time

An initial ANOVA on RT was conducted in which Visual Stimulus (face: Experiment 1A; label: Experiment 1B) was a between-subjects factor along with the within-subject factors of Display Type and FEE. There was a significant main effect of FEE, $F(5, 240) = 18.16, p < .001$, which did not interact with Visual Stimulus. As the RT pattern across FEEs was negatively correlated with the pattern for accuracy, $r(4) = -.84, p < .05$, the FEE effect will not be discussed further. More importantly, there was an interaction between Stimulus and Display, $F(2, 96) = 3.58, p < .05$, partial $\eta^2 = .07$. Pairwise LSD comparisons between face and label stimuli (mean RT = 10.3 and 8.6 ms, respectively) found no statistically significant differences in RT for any of the three display types.

Separate ANOVAs on FEE and Display within each stimulus category revealed that the Display effect was significant only for the face stimuli, $F(2, 48) = 4.30, p = .019$, partial $\eta^2 = .15$. The mean RT with that display was 10.0 s for congruent face displays, 10.1 s for the noise control, and 10.9 s for incongruent faces. Pairwise LSD comparisons found that only the congruent and incongruent displays differed significantly, although the difference between incongruent and noise displays approached significance, $p = .06$.

Influence of visual FEE on distribution of error responses

As was noted in the Introduction, a visual FEE can produce not only a congruence effect (congruent accuracy minus control accuracy) and an incongruence effect (control accuracy minus incongruent accuracy) but also a tendency for errors on incongruent trials to match the category of the visual FEE presented, relative to what would be expected by chance. This *visual-match tendency* was computed as a measure denoted m . The value of m was the proportion of incongruent, incorrect trials where the visual FEE was chosen. That is, for a given haptic FEE, m was the number of incongruent trials where the visual FEE was chosen as a response, divided by the total number of incongruent trials where errors were made (confusions). This analysis pooled data from all visual FEEs associated with the haptic FEE. If the erroneous response was chosen by chance, given that there were 5 potential wrong responses, the expected proportion of matches to the visual FEE would be .20. (Response biases differed across the FEEs, but as this analysis pooled over the responses given to a haptic FEE, the average remained .2.) Figure 3 shows the magnitude of m induced by faces (Experiment 1A) and words (Experiment 1B) for each FEE. Averaged over FEEs, the value of m was .38 for faces and .24 for words, relative to the chance expectation of .20.

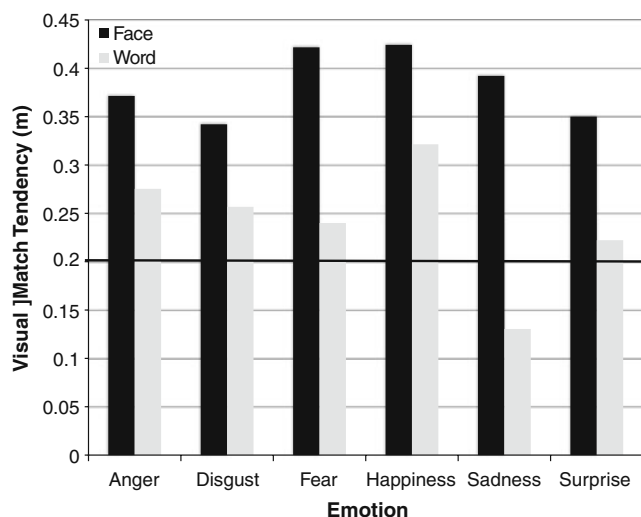


Fig. 3 Visual-match tendency (m) computed from incongruent condition errors for faces and words, by FEE. The horizontal line marks the chance level of .20

Proportion tests comparing the value of m to the chance level of .20, with alpha set to .05, showed that m was significantly greater than chance for all 6 haptic FEEs in Experiment 1A, where photographs of visual faces were presented. In contrast, m was not significantly greater than chance for any haptic FEE in Experiment 1B, where words were used to represent the visual FEEs. Moreover, Fig. 3 shows that for all haptic FEEs, the visual-match tendency m was greater when the visual FEEs were faces than when they were words. When these effects were statistically evaluated with proportion tests comparing the value of m between faces and words, there were significant differences ($p < .05$) for anger, disgust, fear, and sadness. The exceptions for happiness and surprise reflect the low number of errors made for these FEEs, and hence the relative insensitivity of the test, rather than particularly low match tendencies.

Partitioning effects of visual faces

Given that the effects of visual congruence/incongruence on accuracy were significant only for faces, a further analysis focused on the effects in Experiment 1A, using the mean data across FEEs. The goal was to attempt to partition the congruence and incongruence effects of a visual face into two components: a bias to shift categorical assignment toward the visual FEE, and any additional effect that might be present, presumably reflecting additional sources of visual influence. In essence, the analysis asked whether the observed shift in response choices on incongruent error trials could account for the observed level of accuracy on either incongruent or congruent trials, relative to the control. These measures are empirically independent, as a large shift in incongruent errors toward the visual FEE

could occur, whether the actual proportion of errors on incongruent (or congruent) trials was small or large.

The visual-match tendency m , which measures the proportional extent to which an incongruent visual face shifts responses on error trials, was taken as a hallmark of the bias toward the visual category. Accordingly, the analysis first used the value of m , together with the control accuracy, to estimate the magnitude of this bias. This estimate, in turn, was used to predict the accuracy on incongruent and congruent trials, and hence the incongruence and congruence effects (i.e., deviation of incongruent and congruent accuracy, respectively, from the control accuracy). To the extent that these effects were under-predicted by visual bias, it would support the idea that the visual FEE has an effect beyond shifting responses toward its category. The quantitative analysis presented here is intended as a heuristic, since it adopts a very simplistic mechanism for visual influence. Specifically, the analysis assumes that the visual face equally biases two possible outcomes of haptic facemask processing, correct FEE classification and guessing. It can be shown that, under this assumption, the effect of the visual FEE is to modify accuracy relative to the control condition (visual noise). A congruent visual FEE augments control accuracy by converting a proportion of the incorrect trials into accurate responses, and an incongruent FEE reduces control accuracy by converting the same proportion of the accurate trials into errors.¹ The critical proportion is the measure of the visual-category bias.

Note that, according to the analysis, if control accuracy and error rates are approximately equal, this "rob-from-one, give-to-another" adjustment will cause the congruence and incongruence effects to be symmetric. Given that the present control accuracy was close to .5, the asymmetry of the presently observed effects (almost a 2:1 ratio of congruence to incongruence) is an indication that a visually induced category shift, as derived by this approach, cannot entirely account for the effects of the visual FEE on haptic classification responses.

This idea was supported by the quantitative analysis. The visual-category bias, as derived from the mean visual-match tendency m and control accuracy, was estimated as .13. Given the control accuracy of .53, the predicted proportion correct on incongruent trials was then .46. This fell within the 95% confidence interval around the mean accuracy for the incongruent condition ($.44 \pm .04$), indicating that the decrease in proportion correct on incongruent trials relative

¹ It is assumed that, on incongruent trials, visual-category bias produces a shift in trials that would otherwise be correct and a shift in guesses, moving the responses by some proportion toward the FEE associated with the visual face. Bias acts the same on congruent trials, but only shifts in guesses affect performance. As a result, for a visual bias p , congruent accuracy = control accuracy + $p \times$ control error, and incongruent accuracy = control accuracy - $p \times$ control accuracy. A full description of the analysis is available from the first author.

to the control (.09) could entirely be accounted for by the same category bias that shifted error responses toward the visual FEE. In contrast, when visual-category bias was applied to congruent trials, the predicted accuracy was .59, which fell below the lower limit of the 95% confidence interval around the mean accuracy in the congruent condition ($.69 \pm .06$) and led to a substantial underestimate of the advantage for congruent trials relative to the control (.06 predicted vs .16 observed). This underprediction of the congruence effect suggests that it may reflect some mechanism beyond a general bias that shifts responses toward the visually portrayed category.

Discussion

The present study introduces a novel methodology for studying cross-modal interaction in face processing, by manipulating the category-level match between a haptically classified FEE and a simultaneously viewed, but task-irrelevant, visual FEE. The present findings expand on previous conclusions from the literature indicating that haptic/visual interactions in face processing are limited, and when available, vision tends to dominate (Casey & Newell, 2007; Dopjans et al., 2009). Our data indicate that a visual face portraying an FEE has a substantial influence on the classification of the FEE from a simultaneous haptically encoded 3D facemask, as portrayed by another person. Relative to the control condition, where visual noise was presented during haptic FEE identification, participants were 30% more accurate on trials where the haptic and visual faces were congruent and 17% less accurate on trials where they were incongruent. Influence of a visual face was further observed from error trials where the haptic and visual FEEs were incongruent, where there was a clear shift in the distribution of the incorrect responses toward the visual FEE.

In contrast, when visual FEE labels, rather than faces, accompanied haptic FEE identification, there was no difference in accuracy relative to a visual-noise control. In further contrast to the effect of a visual FEE, a visual emotion label did not significantly shift the distribution of error responses to haptic facemasks.

An initial question to consider is whether the effect of a visual FEE on haptic FEE processing might arise at a peripheral level. Specifically, the visual face could influence the pattern of haptic exploration. In this regard, the response-time data indicated that the visual FEEs increased the duration of haptic exploration for incongruent faces relative to congruent ones and the control, suggesting that participants were sensitive to the featural incompatibilities between the modalities and adjusted exploration accordingly. However, congruent visual faces produced no corresponding advantage in speed of exploration, undermining a peripheral account of

the facilitation from matching visual and haptic FEEs. Moreover, even if the visual face affects exploration, that does not offer an explanation of how this translates into an effect on FEE responses. For accuracy to be affected, the information from the visual face must do more than guide haptic exploration; it must be taken into account when the haptic emotion is classified. This implicates mechanisms of interaction at more central levels.

Cross-modal interactions have been hypothesized to result from central processes that combine the inputs from sensory channels. For example, the perceived height of a raised section in a plane can incorporate stereo cues from vision and force cues from touch (Ernst & Banks, 2002). Considering other modalities, flavors paired with sucrose, which are perceived as retronasal odors, become sweeter, whereas those paired with citric acid solutions become sourer (Stevenson, Prescott, & Boakes, 1995). The well-known McGurk effect (McGurk & MacDonald, 1976) shows that visual speech can influence aural recognition of simultaneous phonemes, in some cases to the point of dominance. Conversely, the perceived numerosity of visual displays is changed by co-presented auditory beeps (Shams, Kamitani, & Shimojo, 2000) or tactual taps (Bresciani, Dammeier, & Ernst, 2006). A recent version of the McGurk effect shows that cutaneous air puffs alter the perception of auditory phonemes as aspirated or unaspirated (Gick & Derrick, 2009).

A general rule for such interactions is that the more reliable or precise of two senses, or the less ambiguous, tends to be given greater weight in determining the perceptual outcome, with total capture of one modality by another being the extreme. Predictions of quantitative models based on weighting by information quality, such as the maximum-likelihood model (Ernst & Banks, 2002) and the fuzzy-logical model of perception (Massaro & Friedman, 1990), have been tested and confirmed. Under this approach, it is not surprising that the visual FEE exerts strong influence on haptic identification of facial emotion, given the superiority of visual face processing relative to haptic (as reviewed in Lederman et al., 2010). Moreover, the present data confirm that the intrinsic information content is greater for the visual photographs than the haptically explored facemasks (83% accuracy for classifying FEEs from photographs vs 52% accuracy for the haptic-only control condition), supporting the idea that the visual stimulus will be utilized because it is unambiguous and/or reliable.

It should be noted, however, that the relative weighting of modalities by information quality has most often been assessed with discrepancy paradigms, where each modality receives different information about a particular stimulus. As was indicated in the Introduction, the typical paradigm induces the observer to attribute all the modal inputs to the same physical object. It is neither entirely straightforward nor even appropriate to apply the discrepancy approach to the

present task, which differed from the usual paradigm by offering no hint that the visual and haptic faces stemmed from the same source: they arose from different individuals in different locations, the classification was explicitly focused on the haptic stimulus, and the probability of visual/haptic FEE congruence was deliberately kept low (30%). In short, no matter how clear or reliable the visual channel might be, the present design did not deliberately motivate a subject to use its data. Further relevant is the finding that those FEEs that were identified more accurately by vision, and hence, where the visual input was presumably more informative, did not produce greater congruence or incongruence effects. For example, the visual FEEs for fear and sadness showed low and high labeling accuracy, respectively, but produced congruence effects of comparable magnitude. Moreover, the FEE labels, which might be taken as an extreme of identifiability, had no effect at all. Thus, although an effect of vision even when it is task-irrelevant is generally consistent with the idea of relative weighting by information quality, the present results clearly extend the domain of visual influence.

Although further research is needed to specify more precisely the nature of the vision/touch interaction in the present task, the present results provide some insights into mechanism. One is that the effect of the visual FEE is attributable to the commonality or discrepancy between the haptic and visual faces at a categorical level, where featural invariants of the emotional expression are processed. The portrayal of the simultaneous FEEs by different persons rules out a match at the level of precise physical geometry. The null effect of a label display further indicates that the effects found with visual faces reflect processing of stimuli as faces per se, not just emotional concepts. Further evidence against concept-level influence is the finding that, across emotion categories, there was little relation between the magnitude of the effect produced by a visual FEE in the experiment proper and the accuracy level with which the corresponding emotion was labeled in the preliminary vision-only study.

The present data further offer at least preliminary indications that multiple mechanisms might underlie the influence of an irrelevant visual FEE on haptic FEE classification. A heuristic analysis was offered here as an initial effort to assess whether effects beyond a visually induced category shift might operate in the present task. The tendency to respond with the visual FEE category, as estimated from the error-match measure on incongruent haptic/visual trials, was used to predict the deviations of incongruent-FEE and congruent-FEE accuracy from the control (visual noise). The data showed that the congruence effect was under-predicted. This result reinforces the possibility that something beyond a visual-category bias is operative when people simultaneously view and feel faces making the same FEE.

The present results raise additional intriguing issues about how irrelevant visual inputs might penetrate haptic

FEE classification. One concerns the underlying neural processes that might mediate visual influence. For example, exposure of the visual FEE could dominate cortical circuits that have been identified as contributors to emotional processing shared by vision and touch (Kitada et al., 2010). Interactions at other face processing loci, such as the fusiform face area, could also occur. Another issue concerns the role of visual imagery in the present task. The failure of the emotion label to alter FEE categorization of haptic faces suggests that top-down visual mediation plays little role in haptic processing of FEEs. It is possible, however, that otherwise beneficial visual imagery that would arise bottom-up from haptic exploration is disrupted when an incongruent face is simultaneously seen.

Extensions of the present paradigm may provide a useful tool for addressing these points and further illuminating mechanisms of interaction. One variable that could be explored is the information content in the visual and haptic stimulus, which presumably would affect their reliability-based relative weighting. Information content could be manipulated by varying the type of visual stimulus, for example. As reported above, accuracy of classifying FEEs by vision was found to be higher for the present full-face photographs (83%) than facemasks (72%), for which accuracy was still somewhat higher than visual recognition of raised-line drawings (69% after practice; as reported by Lederman et al., 2008). If reliance on the visual stimulus varies with its information content, as reflected in FEE classification accuracy, one would expect the present congruence and incongruence effects to decline across these categories. Conversely, the influence of vision would be expected to increase if the haptic stimulus was less informative, as raised-line drawings would be in comparison to facemasks. Another useful task would be to compare visual and haptic displays that have geometric congruence to matching at the categorical level only. For example, the visual stimulus could depict the same facemask as the haptic stimulus or a facemask depicting the same FEE expressed by a different actor. Presumably, the addition of physical matching would enhance congruence effects beyond those observed from category-level matches alone.

Acknowledgments This research was supported by NIH Grant 1R01EY016817-01A1 (R.L.K.) and CIHR grant MOP 74645 (S.J.L.).

References

- Arabzadeh, E., Clifford, C. W. G., & Harris, J. A. (2008). Vision merges with touch in a purely tactile discrimination. *Psychological Science*, *19*, 635–641. doi:10.1111/j.1467-9280.2008.02134.x
- Baron, M. (2008). Haptic classification of facial expressions of emotion on 3d facemasks and the absence of a haptic face-

- inversion effect. Unpublished undergraduate honours thesis, Queen's University, Kingston, Ontario
- Bresciani, J., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 554–564. doi:10.1167/6.5.2
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 5497–5502. doi:10.1073/pnas.0935845100
- Casey, S. J., & Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*, *45*(3), 506–513. doi:10.1016/j.neuropsychologia.2006.02.011
- Darwin, C. (1872/1955). *The expression of emotions in man and animals*. New York: Philosophical Library.
- de Gelder, B., Vrommen, J., & Pourtois, G. (1999). Seeing cries and hearing smiles: Crossmodal perception of emotional expressions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*, pp. 425–438. Amsterdam, Netherlands: Elsevier. doi:10.1016/S0166-4115(99)80040-5
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117. doi:10.1037/0096-3445.115.2.107
- Direnfeld, E. (2007). *Haptic classification of live facial expressions of emotion: a face-inversion effect*. (Unpublished undergraduate honours thesis). Queen's University, Kingston, Ontario.
- Dopjans, L., Wallraven, C., & Bühlhoff, H. H. (2009). Cross-modal transfer in visual and haptic face recognition. *IEEE Transactions on Haptics*, *2*, 236–240. doi:10.1109/TOH.2009.18
- Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognising emotions from facial clues*. Englewood Cliffs, NJ: Prentice Hall.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System*. Salt Lake City: UT: A Human Face.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgment of facial expression of emotions. *Journal of Personality and Social Psychology*, *53*, 712–717. doi:10.1037/0022-3514.53.4.712
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. doi:10.1038/415429a. Letter.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*, 502–504. doi:10.1038/nature08572. Letter.
- Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual–haptic integration. *Perception*, *36*, 1523–1533. doi:10.1068/p5851
- Heyman, J. W. (1996). Sensory dominance in infant perception of dynamic expressions of emotion. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(9-B), 5195.
- Kennett, S., Taylor-Clarke, M., & Haggard, P. (2001). Noninformative vision improves the spatial resolution of touch in humans. *Current Biology*, *11*, 1188–1191.
- Kilgour, A., & Lederman, S. J. (2002). Face recognition by hand. *Perception & Psychophysics*, *64*(3), 339–352.
- Kilgour, A., & Lederman, S. J. (2006). A haptic face-inversion effect. *Perception*, *35*, 921–931. doi:10.1068/p5341
- Kitada, R., Johnsrude, I., Kochiyama, T., & Lederman, S. J. (2009). Functional specialization and convergence in the occipitotemporal cortex supporting haptic and visual identification of human faces and body parts: An fMRI study. *Journal of Cognitive Neuroscience*, *21*(10), 2027–2045. doi:10.1162/jocn.2009.21115
- Kitada, R., Johnsrude, I., Kochiyama, T., & Lederman, S. J. (2010). Brain networks involved in haptic and visual identification of facial expressions of emotion: An fMRI study. *Neuroimage*, *49*(2), 1677–1689. doi:10.1016/j.neuroimage.2009.09.014
- Klatzky, R. L., Direnfeld, E., Baron, M., Hamilton, C., & Lederman, S. J. (2010). *Haptic classification of emotional expressions in live faces and 3d facemasks: Efficacy and the role of configural processing*. Manuscript submitted for publication.
- Lederman, S. J., Klatzky, R. L., Abramowicz, A., Salsman, K., Kitada, R., & Hamilton, C. (2007). Haptic recognition of static and dynamic expressions of emotion in the live face. *Psychological Science*, *18*(2), 158–164. doi:10.1111/j.1467-9280.2007.01866.x
- Lederman, S. J., Klatzky, R. L., & Kitada, R. (2010). Haptic face processing and its relation to vision. In J. Kaiser & M. Naumer (Eds.), *Multisensory object perception in the primate brain*. New York: Springer.
- Lederman, S. J., Klatzky, R. L., Rennert-May, E., Lee, J. H., Ng, K., & Hamilton, C. (2008). Haptic processing of facial expressions of emotion in 2-D raised-line drawings. *IEEE Transactions on Haptics*, *1*, 27–38. doi:10.1109/TOH.2008.3
- Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*(2), 215–221.
- Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225–252.
- McGurk, H., & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. doi:10.1038/264746aO. Letter.
- Montgomery, K., & Haxby, J. V. (2008). Mirror neuron system differentially activated by facial expressions and social hand gestures: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*(10), 1866–1877. doi:10.1162/jocn.2008.20127
- Newport, R., Rabb, B., & Jackson, S. R. (2002). Noninformative vision improves haptic spatial perception. *Current Biology*, *12*, 1661–1664. doi:10.1016/S0960-9822(02)01178-8
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5658–5663. doi:10.1073/pnas0400.707.101
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*, 594–596. doi:10.1126/science.143.3606.594
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788. doi:10.1038/35048669. Brief Communications.
- Stevenson, R. J., Prescott, J., & Boakes, R. A. (1995). The acquisition of taste properties by odors. *Learning and Motivation*, *26*, 433–455. doi:10.1016/S0023-9690(05)80006-2
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667. doi:10.1037/0033-2909.88.3.638
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145. doi:10.1037/h0027474
- Zupan, B. (2008). Bimodal perception of visual and auditory cues of emotion from early childhood to early adulthood. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *69*(4-B), 2289.