This article was downloaded by: *[University of Bielefeld]*
On: *30 April 2010*
Access details: *Access Details: [subscription number 917681425]*
Publisher *Psychology Press*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Tone-affect compatibility with affective stimuli and affective responses

Gernot Horstmann [a]
[a] Bielefeld University, Bielefeld, Germany

First published on: 29 April 2010

To cite this Article Horstmann, Gernot(2010) 'Tone-affect compatibility with affective stimuli and affective responses', The Quarterly Journal of Experimental Psychology,, First published on: 29 April 2010 (iFirst)

To link to this Article: DOI: 10.1080/17470211003687538

URL: http://dx.doi.org/10.1080/17470211003687538

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
i*First*, 1–12

Psychology Press
Taylor & Francis Group

Tone-affect compatibility with affective stimuli and affective responses

Gernot Horstmann
Bielefeld University, Bielefeld, Germany

Three experiments examine multimodal integration of tone pitch (high/low), facial expression stimuli (happy/angry), and responses (happy/angry) in a compatibility paradigm. When the participants' task is to imitate facial expressions (Experiment 1), smiles are facilitated by high tones whereas frowns are facilitated by low tones. Experiments 2 and 3 further analyse this effect and show that there is both integration between the tone stimulus and the facial stimulus and between the tone stimulus and the facial response. Results suggest that pitch height is associated with emotion. An interpretation in terms of an embodied cognition approach that emphasizes an interweavement of perception and action is discussed.

Keywords: Emotion; Facial expression; Action; Perception; Embodiment

When presented with multimodal affective stimuli—for example, emotional faces and emotional voices—and asked to judge the conveyed emotion, people readily integrate the information from the two sources (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). This integration is partly involuntary—when participants are instructed to ignore one of the sources (voice or face), its influence diminishes yet does not disappear; the presence of the residual influence of the ignored source on the emotion judgements indicates that intermodal integration is to some degree mandatory (de Gelder & Vroomen, 2000). Corresponding effects with body posture and voice (Van den Stock, Righart, & de Gelder, 2007) and with two visual channels (body posture and facial expression, see van den Stock et al., 2007; face and hand gestures, see Hietanen & Leppänen, 2008) have also been reported. Pourtois, de Gelder, Vroomen, Rossion, and Crommelinck (2000) substantiated that these effects are partly due to changes in perceptual processing: Auditory evoked potentials for angry versus sad voices differed as soon as 120 ms post stimulus onset, depending on whether a simultaneously presented face was emotionally congruent or not. Hietanen, Leppänen, Illi, and Surakka (2004) arrived at the same conclusion with a different approach: They showed that facial expression had no effect on judged vocal emotion when the face was presented 1,500 or 2,000 ms after the voice. To summarize, research revealed that affective information from different modalities is integrated automatically and early in processing.

Correspondence should be addressed to Gernot Horstmann, Psychology Department, Bielefeld University, PO-Box 100 131, 33 501 Bielefeld, Germany. E-mail: gernot.horstmann@uni-bielefeld.de

I am indebted to Rebecca Förster, Sabrina Lehmann, Mirijam Hass, and Lea Rotte for collecting the data and for assessing the latencies for the facial responses.

DOI:10.1080/17470211003687538

In addition to this sensory–sensory integration there is also evidence for an integration of affective sensory and motor representations. Hennenlotter et al. (2005) have shown that perception and expression of facial emotion engage common neural networks in the motor and somato-sensory and limbic-sensory systems of the brain. This indicates that, in addition to the "early" integration between affective signals from different modalities and sources, there is probably also "late"—that is, more action-related—integration of sensory input and motor output. Dimberg and colleagues found evidence for sensory-motor integration with behavioural measures: For example, Dimberg, Thunberg, and Grunedal (2002) revealed faster facial responses (e.g., frowning) in response to compatible stimuli (e.g., an angry face) than in response to incompatible stimuli (e.g., a friendly face). In addition to these visual-motor effects, a study by Warren et al. (2006) found evidence that listening to nonverbal emotional vocalizations modulates premotor activity involved in the control of voluntary facial movements. Results presented by Wallbott (1991) and Niedenthal, Brauer, Halberstadt, and Innes-Ker (2001; see also Oberman, Winkielman, & Ramachandran, 2007) additionally suggest that appropriate motor responses (in particular, facial expressions) may in turn facilitate the perception of congruent affective stimuli (i.e., motor mimicry facilitates the perception of facial expressions). To summarize, evidence suggests that affective sensory information is integrated not only with other sensory information, but also with motor representations.

Research on multimodal integration of affective information has sensibly focused on ecologically valid and complex stimuli such as vocal expressions or facial and bodily gestures. This raises the question whether integration occurs also for much simpler stimulus features that correlate with emotion expression categories. There is suggestive evidence that this might be the case for pitch height. As already reported, de Gelder and colleagues (e.g., de Gelder & Vroomen, 2000) have shown that vocal expressions of emotion are integrated with facial expressions of emotion. This might partly be due to differences in the fundamental frequency (F0) in vocal expressions of different emotions. F0 frequency is an important indicator of vocal emotion with a high F0 frequency characterizing, in particular, happiness (e.g., Juslin & Laukka, 2003; although Scherer, Banse, Wallbott, & Goldeck, 1991, found F0 higher for happiness than for anger only for female, but not for male, voices). Drahota, Costall, and Reddy (2008) found that participants can "hear" smiles versus nonsmiles and based their judgements partly on pitch height. A quite different line of research that connects tone pitch directly to friendliness versus threat comes from ethological research. Morton (1994) reviewed 54 species and found three features characteristic of aggressive/friendly signalling: (a) low-pitched sounds are aggressive, while high-pitched sounds are friendly or submissive; (b) falling tones are aggressive, while rising tones are friendly; (c) rough, atonal sounds are aggressive and soft, melodic sounds are friendly. The rationale for this universalism of the high pitch–friendliness association might be that sound frequency correlates with size, with low voices indicating large size, power, and potential threat and high voices indicating small size and helplessness. Developmental research points in the same direction: Newborns smile preferentially when talked to in a high rather than a low voice (Wolff, 1987), and five-month-olds show positive facial expressions when praised in a high voice and negative facial expressions when scolded in a low voice (Fernald, 1993). Thus, it is possible that the association between emotion and tone pitch is rather due to universal contingencies between certain facial expressions and the pitch of voice. Finally, in musical symbolism, it is well known that high tones are associated with a sprightly humorous mood, while low tones are associated with sad or vigorous majestic moods (Hevner, 1937). Huron, Kinney, and Precoda (2006), following up the ethological work of Morton (1977, 1994), also found that musical pieces played in higher pitch are more submissive and less threatening than when played in a lower pitch. To summarize, pitch height appears to be an important affective marker in nonverbal affective signalling.

The present experiments

The present experiments test the hypothesized intermodal integration between tone pitch and facial expressions of emotion. They are intended to extend our knowledge in four respects.

First, previous research on auditory–visual integration of affective information has dealt with natural vocal expressions of emotion, which are multidimensional and complex in patterning over time. It is thus of interest whether integration also occurs with more isolated dimensions, such as sound frequency, as has been implicated in affective prosody and nonverbal signalling. This is tested in the present study.

Second, while previous research on sensory–sensory integration (e.g., de Gelder & Vroomen, 2000) has shown that visual and auditory indicators of emotion are integrated, it is unknown whether a corresponding integration also occurs with affective motor responses. The present study tests both the integration of an auditory stimulus (a high vs. a low tone) with the visual stimulus of a facial expression of emotion (happy vs. angry) and with the actual execution of a facial expression of emotion (happy vs. angry).

Third, while multimodal integration has previously been observed in tasks that require an explicit categorization of the affective stimuli, the present Experiment 1 uses a facial imitation task that only implicitly involves the categorization of the affective stimulus. The finding of multimodal integration under these conditions would thus corroborate de Gelder and Vroomen's (2000) hypothesis of the mandatory and implicit nature of multimodal integration.

Fourth, while affective sensory–motor interactions have been the target of some research (e.g., Niedenthal et al., 2001), former research used customized tasks rather than established paradigms from the mental chronometry approach that is typical for cognitive psychology. The present experiments are conducted using a compatibility task, where intermodal integration effects register in response time (RT) differences between congruent and incongruent stimulus–stimulus or stimulus–response pairings. This methodological step facilitates the communication and comparison of results with research communities outside multimodal integration research.

EXPERIMENT 1

Method

Participants
A total of 12 volunteers (9 women), with a mean age of 26.3 years ($SD = 5.2$ years) participated in exchange for candy or course credits. Three additional cases were discarded—two (both women) because they had difficulties in discriminating the tones and one (a man) because he accidentally moved the mirror used to reflect the stimuli on the videotape (explained below).

Stimuli
The high (600-Hz) and the low (300-Hz) tone were presented with a duration of 100 ms. The pictures were 12 photos (256-color) from the JACFEE (Japanese and Caucasian Facial Expressions of Emotion; Matsumoto & Ekman, 1988) displaying a happy (3 women and 3 men) or an angry face (3 women and 3 men). The faces covered the entire screen of a 15″ monitor and were thus almost the size of real faces.

Apparatus
ERTS (Experimental Run Time System), run on a Pentium II PC, controlled stimulus presentation and manual response registration. Facial responses were recorded with a SONY camcorder, mounted on a tripod behind the monitor. A mirror on a tripod, positioned behind the participant, reflected the stimulus from the computer screen to the camera. The camera was positioned such that both the facial responses and the mirror reflection of the stimulus were combined. The video was analysed offline with Adobe Premiere.

Procedure
Participants were instructed to mirror the facial expression presented on the monitor. Speed and accuracy were emphasized for this task. The

expression was to be held until the picture disappeared. The instructions also informed about the concomitant tone, which had to be responded to after the picture. Participants were to respond with either the left or the right response key (balanced across participants) to the high versus low tone, respectively. The tone response was nonspeeded; only accuracy was emphasized.

Participants worked through five blocks of 26 trials each. Each block began with 2 warm-up trials (not analysed), randomly drawn from the four experimental conditions (see below), followed by 24 experimental trials.

Each trial began with the message "Neutral face!" for 2 s, followed by a fixation cross for 1 s. Next the tone and the picture started simultaneously. The tone lasted for 100 ms and the picture for 1,500 ms. After the picture, the screen was blank for a maximum of 8 s. A manual response terminated the blank screen.

Design

Each block consisted of two presentations of each photo: one where the photo was combined with a high tone and one where the photo was combined with a low tone.

Analysis

The main dependent variable was the latency of the speeded facial response. The video recordings were analysed with Adobe Premiere editing software for digital films. The facial response's onset in each trial was judged by playing back and forth the mute video recording of each trial's face response until the frame where the movement of the correct response started was found. Happy and angry faces are characterized by the movement of different muscle groups—that is, the smile (mostly zygomaticus major activity) in the happy face and the constrictions of the brows (corrugator and depressor supercilii) and pressed lips (orbicularis oris) in the angry face—and there is also some variation between participants in the production of the faces. Thus, it was decided that the onset of any movement in the mouth or in the brow area that preceded the correct facial expressions was taken as the onset of the movement. (Technically, the task may be described as a perceptual judgement of movement onset.) If no correct face was presented during the 2,000-ms presentation of the facial stimulus, the response was coded as an error. In each trial with a correct response, the ordinal number of the frame (relative to the starting frame 0, which was the onset of the facial picture) was recorded. For the coding, the video tone was turned off, such that the raters were blind with respect to the congruency conditions. Facial RTs in milliseconds were computed as the number of frames multiplied with the duration of one frame (40 ms). The temporal resolution of the RT measurement might seem coarse on first sight, but high-resolution measures can be obtained on a low-resolution system by averaging repeated measures. Course sampling can be viewed as adding random errors to the true value, which are cancelled by averaging over repeated measures.

Results

Facial RTs shorter than 200 ms or longer than 1,500 ms were discarded. Mean correct facial RTs (see Figure 1a) were analysed with a 2 (tone: high versus low) × 2 (facial emotion: happy versus angry) analysis of variance (ANOVA), which revealed a significant Tone × Facial Emotion interaction only, $F(1, 11) = 23.97$, $p < .001$ (main effects Fs < 1.86). Happy faces were produced faster when presented with high than with low tones (578 vs. 661 ms), $t(11) = 4.1$, $p < .01$, and angry faces were produced faster when presented with low than with high tones (633 vs. 585 ms), $t(11) = 2.9$, $p < .05$. Figure 1a shows the main results.[1] Analysis of errors was not conducted because these were too infrequent (<1%).

---

[1] Visual inspection of the data revealed that the results were the same for male and female participants, which is also revealed by a corresponding ANOVA with sex of participant as an additional variable (all Fs < 1). Note, however, that there were only 3 men in the sample.

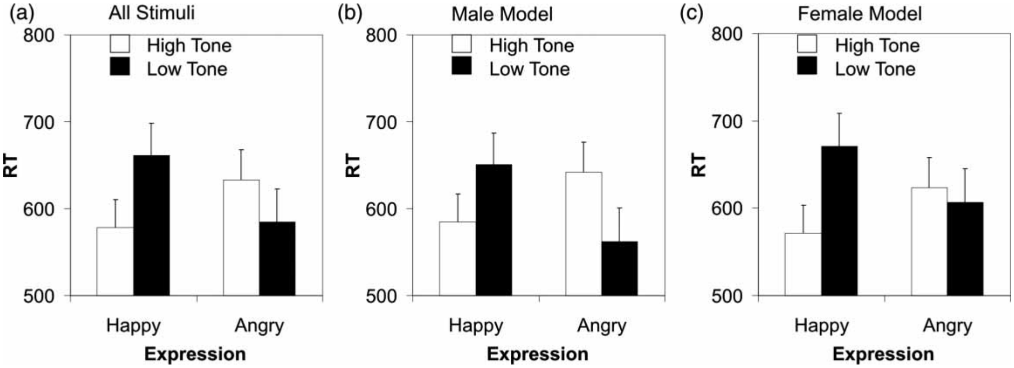

Figure 1. *Results for Experiment 1: Facial response times (RTs) for (a) all stimuli, and for only the (b) male, and (c) female models. Error bars show the standard error of the mean.*

Because displayer gender might interact with the compatibility effect, the means were recomputed, and displayer gender was added as a variable. In addition to the already reported Tone × Facial Emotion interaction, the triple interaction was in fact significant, $F(1, 11) = 11.64$, $p < .01$. Figures 1b and 1c show the results for male and female models, respectively. To clarify this triple interaction, ANOVAs with the variables tone and displayer gender were conducted separately for happy and angry faces. For happy faces, this post hoc analysis revealed a tone effect only, but no interaction with sex of displayer: Tone × Sex of displayer interaction, $F(1, 11) = 2.07$, $p = .17$. In contrast, for angry faces, the tone effect (difference between high and low tones) was larger with male displayers (80 ms) than with female displayers (16 ms), $F(1, 11) = 6.37$, $p < .05$. A possible explanation for this unanticipated result is that gender adds an additional congruency dimension into the equation, as women are less associated with anger than men (cf. Hess, Adams, & Kleck, 2007) and as women are arguably more associated with high pitch than men (because of sex differences in F0 frequency).

Although not directly relevant for the present research question, the manual responses for the tone discrimination task were also analysed for completeness. The instructions required accurate, but not speeded, responses; thus the analysis was conducted on the accuracy scores (rather than RTs). An ANOVA with the same variables as before revealed no main effects, $F\text{s} < 1$, but a Facial Emotion × Tone interaction, $F(1, 11) = 5.81$, $p < .05$, revealing more errors in the incongruent (14%) than in the congruent (5%) conditions. It is unclear whether this is an independent effect of facial expression on tone discrimination, or whether it is merely an index of the interference induced in the first task. This question, however, will not be pursued in the present experiments and may be answered by future research.

Discussion

Experiment 1 found a strong congruency effect between tone pitch height and facial expression: A happy face was imitated faster when combined with a high rather than with a low tone, while an angry face was imitated faster when combined with a low rather than with a high tone. This result concurs with the literature on multimodal integration of affective information, which has substantiated that auditory and visual affective information is readily integrated (e.g., de Gelder & Vroomen, 2000; Massaro & Egan, 1996). While previous experiments, however, did use natural stimuli that are multidimensional, the present experiments tested multimodal integration between the isolated feature of pitch height and multidimensional facial expression. Pitch height has been implicated by research from various disciplines as a component of vocal expressions of emotion in men and animals. The present results thus suggest that one key feature of vocal expressions that is integrated with facial expressions of emotion is pitch height.

The task in the present experiment was deliberately designed to resemble natural behaviour (smiling when another person is smiling) rather than the typical laboratory task behaviour (pressing keys to indicate a perception or decision). In particular, identifying the depicted emotion was not explicitly requested by the experimenter, though it was an implicit component of the imitation task. The present results thus reinforce de Gelder and Vroomen's (2000) hypothesis that intermodal integration is a mandatory process. While these authors showed that integration cannot be shut off deliberately (i.e., the irrelevant channel cannot be easily ignored), the present results suggest that integration occurs incidentally. It is interesting to note that the auditory stimulus was the target of a separate, nonemotional task in the present experiment. Yet, participants were apparently unable to process tone and face separately.

While the facial imitation task has the strong advantage of testing multimodal integration in the context of relatively natural behaviour, it has the disadvantage of confounding effects of stimulus–stimulus and stimulus–response associations. This can easily be seen when the experiment is more formally analysed. Participants were confronted with a dual-task situation. The main task, Task 1, was to imitate (Response 1, R1) a happy or angry face shown as a visual stimulus (Stimulus 1, S1). The secondary task, Task 2, which had to be executed after Task 1 had been finished and which was not speeded, was to categorize a low or high tone (S2) with a key press (R2).

According to the most relevant theoretical model here, the dimensional overlap model (DOM; Kornblum, Hasbroucq, & Osman, 1990), congruency effects occur when there is a dimensional overlap among simultaneously presented stimuli or between stimuli and responses. A typical example for a dimensional overlap among stimuli (or different stimulus features of the same object) is a Stroop stimulus—that is, a colour word printed in a colour different from its lexical meaning (e.g., the word "green" printed in red). Naming the colour of a Stroop stimulus is difficult because of the dimensional overlap between the colour name and the lexical meaning of the word (both imply represented colour categories). A straightforward example of a dimensional overlap between a stimulus and a response is the spatial compatibility effect (Broadbent & Gregory, 1962). Here, stimuli and responses vary on a horizontal spatial dimension; left or right stimuli are responded to fast with the left or right hand, respectively, while the reversed mapping is slow. In terms of the DOM, stimulus and response codes overlap in the dimension of horizontal spatial positions, which in turn facilitate responses in the case of congruency and lead to interference in the case of incongruence between stimulus and response codes.

Applied to the present task, the auditory tone stimulus, the visual facial stimulus, and the facial motor response are expected to show compatibility effects to the degree that there is a dimensional overlap between their representations. In particular, a high tone should be compatible with a happy face stimulus and a happy facial response, and a low tone should be compatible with an angry face stimulus and an angry facial response (of course, a happy face response to a happy face picture is also highly compatible, but this compatibility is not directly tested here). Compatible conditions, in turn, should reveal shorter RTs than incompatible conditions, which are, in particular, a high tone paired with an angry face stimulus or response and a low tone paired with a happy face stimulus or response.

Based on this analysis, Experiment 1 leaves open the theoretically important question of whether the processing of the stimuli S1 and S2 interact (an S–S effect) or whether the auditory stimulus interacts with the facial response (an S–R effect). For this reason, Experiment 2 replaced the facial response with an affectively neutral key press to test for S–S effects between the tone and the facial response. Correspondingly, Experiment 3 replaced the facial stimulus with affectively neutral symbols and tested for S–R effects between the tone and the facial response.

EXPERIMENT 2

Experiments 2 and 3 examine whether the effect observed in Experiment 1 is due to an interaction of the two stimuli or to an interaction between the tone and the facial response. Experiment 2 tests whether the presentation of a tone and a facial stimulus, in the absence of a facial response, is sufficient for the congruency effect to show up. To this end, the facial stimuli were responded to by pressing a key under speed instructions, whereas the tone was responded to with an unspeeded vocal response, to be executed after the key press.

Method

Participants
A total of 12 volunteers (10 women), with a mean age of 24.3 ($SD = 4.9$) years took part for course requirements or candy. These had not taken part in Experiment 1.

Stimuli and procedure
In Experiment 2, new stimulus pairs were used. The tones had frequencies of 300 Hz and 1,600 Hz. The facial stimuli were colour photos of a young woman (see Figure 2) showing a happy or an angry face, respectively. Only 2 photos (instead of 12 as in Experiment 1) were used to match conditions of Experiments 2 and 3 with respect to the number of stimuli and responses (viz., Experiment 3 presents two neural stimuli and requires two affective responses, corresponding to the two affective stimuli and the two neutral responses in the present experiment). Variations in stimulus and task parameters are also useful to probe the robustness of the effects.

The procedure was the same as that in Experiment 1 with the exception that the tasks were different: The two faces had to be responded to by pressing one of two response keys (left or right, counterbalanced across participants), and the two tones had to be responded to by saying high or low. As in Experiment 1, Task 1 was to be performed under speed instructions, while Task 2 was to be responded to under accuracy instructions (the experimenter noted the responses on a protocol form). There were 128 experimental trials in total, evenly distributed over the four experimental conditions.

Figure 2. *The stimuli used in Experiment 2 (originals were coloured).*

Results

Manual RTs below 200 ms or exceeding 1,500 ms were discarded. Mean manual RTs (see Figure 3) were analysed with a 2 (tone: high versus low) × 2 (facial emotion: happy versus angry) ANOVA, which revealed a significant Tone × Facial Emotion interaction only, $F(1, 11) = 27.91$, $p < .001$ (other main effect and interactions: $F\text{s} < 1$).

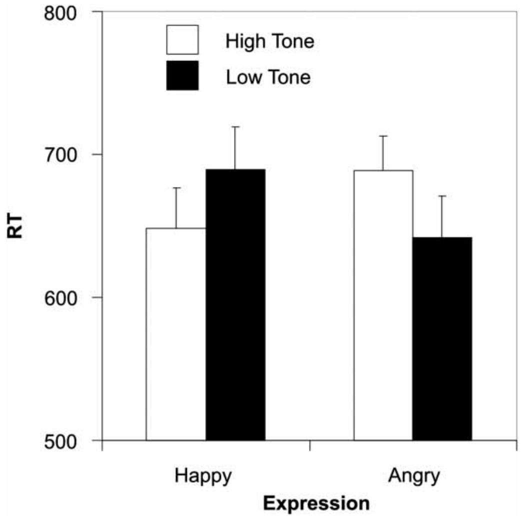

Figure 3. *Manual response times (RTs) for Experiment 2. Error bars show the standard error of the mean.*

Happy faces were responded to 57 ms faster when presented with high than with low tones, $t(11) = 2.37$, $p < .05$, and angry faces were responded to 64 ms faster when presented with low than with high tones, $t(11) = 5.76$, $p < .001$. Error rates were too low (<3 %) to warrant further analysis. Errors in the tone classification task were virtually absent (<1%).

Discussion

Experiment 2 showed a strong congruency effect between tone pitch height and facial expressions when only the stimulus, but not the response, was emotional. In fact, the congruency effect was not markedly weaker in the absence of an emotional response than it was in the presence of an emotional response (Experiment 1). This indicates that large amounts of the effect could be explained by the correspondence between the facial and the auditory stimuli.

Experiment 2 used only one pair of stimuli, with one stimulus only for each emotion category. With only two stimuli, it is possible that participants used superficial perceptual aspects of the pictures to determine their responses rather than emotion. However, because photos of the same person were used, differences other than in facial expression should be minimal. Moreover, even if participants based their response on nonemotional aspects, the unattended affective expressions would still be the most plausible cause of the congruency effect, simultaneously implying that the congruency effect is independent of attention. Finally, Experiment 2 shows the same pattern as that in Experiment 1, where the task required categorization rather than discrimination.

EXPERIMENT 3

Experiment 2 had shown that an intrinsically affective response is not necessary for the congruency effect. Experiment 3, in turn, asks whether an affective response, in the absence of an affective stimulus, is sufficient to instantiate a tone–affect compatibility effect. Therefore, Experiment 3 presented nonemotional stimuli—the @ and the & symbol—and participants were instructed to respond with a happy, or angry, face, respectively (stimulus–response mapping was reversed for half of the participants).

Method

Participants

A total of 12 volunteers (9 women) with a mean age of 27.1 (*SD* = 5.0) years participated in the experiment. These had not participated in any of the previous experiments.

Stimulus and procedure

This was basically the same as that in Experiment 1, except that (a) the stimuli were the symbols @ and &, and (b) the high tone was 1,600 Hz (comparable to Experiment 2).

Results

Facial RTs below 200 ms or exceeding 1,500 ms were discarded. Mean facial RTs (Figure 4) were analysed with a 2 (tone: high vs. low) × 2 (facial emotion: happy vs. angry) ANOVA, which

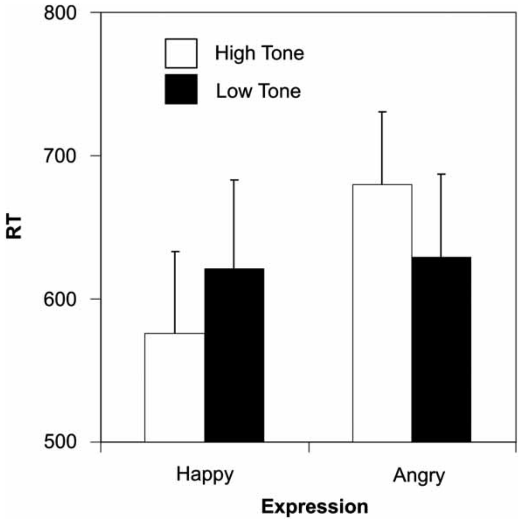

Figure 4. *Facial response times (RTs) for Experiment 3. Error bars show the standard error of the mean.*

revealed a significant Tone × Facial Emotion interaction, $F(1, 11) = 18.46, p < .001$, and a significant main effect of facial emotion, $F(1, 11) = 9.92, p < .01$ (main effect tone $F < 1$). Happy faces were produced 45 ms faster when presented with high than with low tones, $t(11) = 3.32, p < .01$, and angry faces were produced 51 ms faster when presented with low than with high tones, $t(11) = 3.64, p < .01$. Errors were very infrequent and thus were not analysed.

As in Experiment 1, the errors for the tone response were analysed using the same design as that for the facial RTs. The analysis revealed a main effect for facial valence, $F(1, 11) = 6.9, p < .05$, reflecting somewhat fewer errors with happy than with angry faces (4% vs. 6%).

Discussion

Experiment 3 shows that the affective congruency effect between tone and facial emotion can be found when an affective facial response was made to an emotionally meaningless stimulus. Thus, Experiment 3 succeeded in revealing a pure case of affective sensory–motor integration. To summarize, Experiments 2 and 3 together show independent integration effects of tone pitch height and the perception of an affective facial stimulus, on the one hand, and the production of an affective facial response, on the other hand.

GENERAL DISCUSSION

Three experiments found strong affective congruency effects between tone pitch and facial expressions of emotion. Experiment 1 found that facial imitation was influenced by a tone that was uncorrelated with the imitation response: Happy faces were imitated faster with high- than with low-pitched tones, while angry faces were imitated faster with low- than with high-pitched tones. According to a very general model to analyse compatibility effects, the dimensional overlap model (DOM), the results reveal that tone pitch frequency and facial expressions overlap in some dimension. Analysis using the DOM, however, also showed that Experiment 1 leaves open whether the effects are due to an interaction of the two stimuli—that is, the tone and the facial picture (a stimulus–stimulus effect)—or of the tone stimulus and the facial response (a stimulus–response effect). Therefore Experiments 2 and 3 were conducted to isolate stimulus–stimulus and stimulus–response effects. The results showed that both stimulus–stimulus and stimulus–response effects exist.

The results of Experiment 2 converge with the results from previous cross-modal integration experiments (e.g., De Gelder & Vroomen, 2000; Hietanen et al., 2004). These experiments have shown that irrelevant vocal expressions of emotion are integrated with visual expressions of emotion even when the instructions clearly stated that they be ignored. In the present Experiment 2, a short tone nominally belonging to a separate task had corresponding effects on the categorization of affective facial expressions. These effects are predicted on the hypothesis that pitch height is a feature of vocal expressions of happiness in humans (e.g., Juslin & Laukka, 2003) and may even be a universal indicator of friendliness versus threat over species (Morton, 1994).

The results of Experiment 3 extend previous research in that they show that cross-modal integration can be observed not only in the perception of expressions of emotion, but also in the production of expressions of emotions. The production of a happy or an angry face was affected by the task-irrelevant tone in much the same way as the perception of a happy or angry face. This result shows, for the first time, multimodal integration involving the actual execution of an affective response (i.e., affective facial motor behaviour). As is discussed in more detail below, this result fits within an embodiment approach to representation, in this case of emotion categories.

The results from Experiment 3 also connect to the few existing studies examining direct sensation–motor or cognition–motor interactions in social psychology. For example, a frequently cited effects in the past decade has been the slowing of walking speed in student participants induced by the unobtrusive priming of an elderly stereotype (Bargh, Chen, & Burrows, 1996). Correspondingly, Experiment 3 shows the priming of socially relevant behaviour (facial expressions of emotion) by the perception of task-irrelevant tones.

If we accept the DOM as the adequate model for the present effects, what then is the basis of the compatibility effect? Or in other words: If the compatibility effect is due to a dimensional overlap, what then are the overlapping dimensions? As reviewed in the introduction, evidence from research on vocal affect, development, music perception, and cross-species similarities suggests that the auditory representations of happiness probably include the dimension of sound pitch, specifying that happy sounds are (relatively) high, while nonhappy sounds are relatively low. Additionally, the same sources suggest that the auditory representation of anger and dominance includes low-pitched sounds. Thus, the common currency of high- versus low-pitched sounds and happy versus angry facial expressions of emotion appears to be an affective feature.

If this is accepted, the next question naturally arises as to whether the association is direct or mediated. The evidence for multimodal integration (see introduction) suggests that the mental representations of happiness and anger include auditory, visual, motor components, and so on—that is, these representations are multimodal (as are their referents). On this account, hearing affective indicators of emotion (for example happiness) automatically activates other modal (visual, motor) components in the multimodal representation of that emotion. Accordingly, compatible affective responses are facilitated, and incompatible affective responses are impeded.

Arguably, there is a second possibility that the dimensional overlap is not direct, but rather mediated by a further representation that is associated with both the tone and the facial expression. For example, Walker and Smith (1984) report that research on phonetic symbolism suggests that high-pitched vowel sounds are used to represent smallness, sharpness, brightness, quickness, activity, weakness, and angularity. These authors also show that corresponding effects can be unveiled with a compatibility task very similar to the present one, where visually presented words that are accompanied by irrelevant high or low tones have to be categorized. Walker and Smith interpret these effects as suprasensory—that is, based on abstract (conceptual), but not sensory (modal) categories. In support of this notion it might be argued that the tones are conceptually categorized as high or low immediately on presentation, because this was the relevant dimension for the secondary task later in the trial. Similarly, happy and angry faces are conceptually categorized on their appearance, which in turn activates associated concepts, inter alias, the concepts high and low. Research on language and representation strongly suggests that high and low spatial positions are a common metaphor for positive and negative states of affairs (heaven is up, hell is down; Lakoff & Johnson, 1980; Meier & Robinson, 2004). The conceptualization of tones and faces as high or low, then, is the basis for the dimensional overlap.

Neither of these two accounts can be excluded with certainty on the present data. However, the fact that the task in Experiment 1 was not to categorize the facial stimuli as positive or negative but rather to imitate them renders the conceptual account somewhat less plausible. Imitation of facial expressions is almost the prototype for an ideomotor compatible action, where the perception of the stimulus is sufficient to specify the corresponding response. Accordingly, developmental studies show early imitation in infants (Meltzoff & Moore, 1977), electrophysiological studies show motor activation of corresponding muscles during the perception of facial expressions (Dimberg, 1982), and functional brain imaging studies show shared motor activation for the perception and the production of facial expressions (Hennenlotter et al., 2005). Conceptual representations of the stimuli or the responses do not appear to be necessary in a facial imitation task that is so natural for humans.

The present research fits nicely within an embodiment perspective of mental representation. This perspective opposes the more classical semantic network approach that conceptualizes, for instance, emotions as nodes in a propositional semantic network. The classical approach has been very successful in modelling memory related effects such as mood-dependent memory, affective priming, or the influence of affect on creativity (e.g., Bower, 1981; Clark & Isen, 1981) through the concepts of spreading activation. However, because semantic codes are propositional, while perceptions and actions are analogical and modality bound (which are incommensurable representational formats), additional processes are needed to translate sensory representations into propositional ones and to translate propositional representations into actions.

In contrast, the embodiment approach emphasizes that so called "central" cognitive and affective processes are not abstracted from their perceptual or motor origin, but are rather grounded in modal sensory and motor processes (e.g., Barsalou, 2008; Fischer & Zwaan, 2008; Glenberg & Kaschak, 2003; Niedenthal, 2007; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). For example, when thinking about how happy one could be after a certain decision, quite a number of representations from different modalities would be activated, dealing with sensory concomitants of happiness, such as the perception of the happiness-triggering event, and the perception of emotional and bodily feelings as well as with movements and behaviours such as smiling, jumping, and cheering. Thus, the embodiment approach naturally bridges the gap between perception, cognition, and action, enabling the immediate interaction between stimuli from different modalities, as well as interactions of stimuli and responses. Thus, the embodiment approach readily assimilates the present results of multimodal integration both between two stimuli as well as between a stimulus and a response.

To conclude, multimodal affective integration proceeds already on a subcategory level, with tone pitch height having similar effects as had previously been reported for vocal expressions of emotion. Multimodal integration can be observed when no explicit emotion categorization is required by task, but where emotion categorization is only implicitly required in the course of a relatively natural facial imitation task. Similar effects are observed for affective stimuli and affective responses, testifying the generality of the phenomenon, and supporting the embodiment perspective stating that representations, at least of natural categories, are best understood as being multimodal, including sensory as well as motor modalities.

Original manuscript received 16 January 2009
Accepted revision received 22 January 2010
First published online day month year

REFERENCES

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait constructs and stereotype activation on action. *Journal of Personality and Social Psychology*, *71,* 230–244.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, *59,* 617–645.

Bower, G. (1981). Mood and memory. *American Psychologist*, *36,* 129–148.

Broadbent, D. E., & Gregory, M. (1962). Donders' B- and C-reactions and S–R compatibility. *Journal of Experimental Psychology*, *63,* 575–578.

Clark, M. S., & Isen, A. M. (1981). Towards understanding the relationship between feeling states and social behavior. In A. Hastorf & A. M. Isen (Eds.), *Cognitive social psychology* (pp. 73–108). Amsterdam: Elsevier/North-Holland.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14,* 289–311.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, *19,* 643–647.

Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition and Emotion*, *16,* 449–471.

Drahota, A., Costall, A., & Reddy, V. (2008). The vocal communication of different kinds of smile. *Speech Communication*, *51,* 278–287.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Developmental Psychology*, *64,* 657–674.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61,* 825–850.

Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 93–126). New York: Academic Press.

Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., et al. (2005). A common neural basis for receptive and expressive communication of pleasant facial affect. *NeuroImage*, *26,* 581–591.

Hess, U., Adams, R. B., & Kleck, R. E. (2007). When two do the same, it might not mean the same: The perception of emotional expressions shown by men and women. In U. Hess & P. Philippot (Eds.), *Group dynamics and emotional expression* (pp. 33–50). New York: Cambridge University Press.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, *49,* 621–630.

Hietanen, J. K., & Leppänen, J. M. (2008). Judgment of other people's facial expressions of emotions is influenced by their concurrent affective hand movements. *Scandinavian Journal of Psychology*, *49,* 221–230.

Hietanen, J. K., Leppänen, J. M., Illi, M., & Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology*, *16,* 769–790.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review*, *1,* 170–177.

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129,* 770–814.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus response compatibility—a model and taxonomy. *Psychological Review*, *97,* 253–270.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3,* 215–221.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE)* [CD-ROM]. San Francisco: Intercultural and Emotion Research Laboratory.

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

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198,* 75–78.

Morton, E. (1977). On the occurrence and significance of motivation-structural rules in some birds and mammal sounds. *The American Naturalist*, *111,* 855–869.

Morton, E. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge, UK: Cambridge University Press.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316,* 1002–1005.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9,* 184–211.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, A. H. (2001). When did her smile drop? *Cognition and Emotion*, *15,* 853–864.

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). *Social Neuroscience*, *2,* 167–178.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11,* 1329–1333.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, *15,* 123–148.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7,* 487–494.

Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, *13,* 75–81.

Wallbott, H. G. (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology*, *30,* 207–219.

Warren, J. G., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., et al. (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *The Journal of Neuroscience*, *26,* 13067–13975.

Wolff, P. H. (1987). *The development of behavioural states and the expression of emotions in early infancy*. Chicago: University of Chicago Press.