

Tilburg University

## Is cross-modal integration of emotional expressions independent of attentional resources?

Vroomen, J.; Driver, J.; de Gelder, B.

*Published in:*  
Cognitive, Affective and Behavioral Neuroscience

*Publication date:*  
2001

*Document Version*  
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective and Behavioral Neuroscience*, 1(4), 382-387.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

In: **Cognitive, Affective, and Behavioral Neuroscience**, Vol 1 (4), Dec 2001,  
pp382-387.

## Is Cross-modal Integration of Emotional Expressions Independent of Attentional Resources?

Jean Vroomen<sup>1</sup>, Jon Driver<sup>2</sup> and Beatrice de Gelder<sup>1</sup>,

<sup>1</sup>*Tilburg University, the Netherlands*

<sup>2</sup>*University College, London, U.K.*

Running head: Attention and cross-modal emotion

Correspondence address:

Jean Vroomen  
Tilburg University  
Dep. of Psychology  
P.O. box 90153  
5000 LE TILBURG  
The Netherlands  
Phone: +31-13-4662394  
e-mail: j.vroomen@kub.nl

### Abstract

In this study, we examined whether integration of visual and auditory information about emotions requires limited attentional resources. Participants judged whether a voice expressed happiness or fear, while trying to ignore a concurrently presented static facial expression. As an additional task, participants might have to add two numbers together rapidly (Experiment 1), count the occurrences of a target digit in a rapid serial visual presentation (Experiment 2), or judge the pitch of a tone as high or low (Experiment 3). The visible face had an impact on judgments of the heard voice emotion in all experiments. This cross-modal effect was independent of whether or not subjects performed a demanding additional task. This suggests that integration of visual and auditory information about emotions may be a mandatory process, unconstrained by attentional resources.

Many events have perceptual consequences in more than one sensory modality. For instance, a person talking produces both audible speech-sounds and visible lip-movements, and these can have a combined influence on speech perception, as is now well-established (McGurk & MacDonald, 1976). More recently, it has been shown that crossmodal interactions can also influence affective judgements. That is, expressed emotion may also be recognized by a combination of the heard voice and seen face. De Gelder and Vroomen (1995; 2000) observed that subjects judging the emotional tone of a voice (happy or fearful) were systematically affected by the expression of a static face (e.g., the voice was judged as less fearful if the face was smiling). This crossmodal bias was present even when subjects were told to ignore the face and concentrate on the heard voice only (see also Massaro & Egan, 1996).

The fact that this crossmodal influence was observed even when subjects were instructed to ignore the visual modality (i.e. to ignore the visible face) may indicate that crossmodal integration of affective information takes place automatically, regardless of attentional factors. Massaro (1998, pp. 246), though, has shown that instructions to pay selective attention to either the auditory or the visual modality can increase the influence of the modality to which attention is directed. This shows that selective attention to a particular modality can modulates cross-modal effects. From these data, though, it is not clear whether attention is actually required for a cross-modal effect to obtain. For example, if participants had to judge the voice, would any cross-modal effect from the face on voice judgments occur if the face were not attended?

Recent attention research has shown that irrelevant visual stimuli may be particularly hard to ignore under 'low-load' conditions in the prescribed task, yet can be successfully ignored under higher-load conditions, where the specified task consumes more attentional capacity (e.g. Lavie, 1995; 2000). The task-irrelevant face in de Gelder & Vroomen's (2000) study on judgements of heard emotion might therefore have been unusually hard to ignore, due to the low-load nature of the situation (the face was the only visual stimulus present; and the only task requirement was judgement of the voice). It is thus possible that the influence of the seen facial expression on judgements of the emotional tone of a heard voice would

be eliminated under conditions of higher attentional load (e.g. with additional visual stimuli present, and a demanding additional task). On the other hand, if crossmodal integration of affective information is a truly automatic process, it should arise regardless of the demands of any additional task. Indeed, independence from demands on attentional capacity has long been one of the defining characteristics of 'automatic' processes (e.g. see Shiffrin & Schneider, 1977).

In this study we examined by means of a dual-task methodology whether cross-modal integration of affective information suffers when a demanding task has to be performed concurrently. A positive result would suggest that attentional resources are required for cross-modal integration of emotion to arise. If, on the other hand, a competing task does not affect performance, one may assume that cross-modal interactions do not require attentional resources in order to proceed (Kahneman, 1973). In the present study, we measured the influence of a visible face's static expression on judgements of the emotional tone of a heard voice (as in de Gelder & Vroomen, 2000), while varying attentional demands by presenting the subject with an additional task.

A general concern in applying the dual-task methodology is whether tasks compete for the same pool of resources, or whether there are multiple resource pools that each deal separately with the various cognitive and perceptual aspects of the two tasks (Wickens, 1984). When tasks do not interfere, it may be either that one of the tasks (or both) do not require any attentional resources (i.e., they are performed automatically), or instead that they draw on different resource pools. To distinguish these alternatives, we varied the nature of the additional task. If none of the different tasks interferes with the cross-modal interactions, this would suggest that the cross-modal effect itself does not require attentional resources. As additional task, we used an 'adding' task in Experiment 1 (i.e., adding two digits together), a 'visual' task in Experiment 2 (i.e., detecting zeroes in a rapidly presented sequence of digits), and an 'auditory' task in Experiment 3 (i.e., deciding whether a tone was high or low).

## Experiment 1

### Method

Participants. A group of sixteen college students from Tilburg University was

tested. They received course credits for their participation.

Auditory materials. The same materials were used as described in de Gelder and Vroomen (2000, Exp. 3). In brief, the prosody of a Dutch sentence with semantically neutral content (>Zijn vriendin komt met het vliegtuig=, >His girlfriend is coming by plane=) was modified so as to create a continuum between 'happy' and 'fearful' prosody. The continuum was created by changing simultaneously the duration, pitch range, and pitch register of the utterances. In order to change the pitch in equal steps, the original pitch contour was replaced by a minimal sequence of straight-line approximations while the perceptual identity remained close to the original one. A program computed the various pitch movements by superimposing them on a declination line. Then, only two free parameters needed to be set: the excursion size of the pitch movements in semitones and the place in the pitch register. For the 'happy' endpoint of the continuum, the excursion size was set at 10 semitones and the end-frequency at 150 Hz. For each successive stimulus in the continuum, the excursion size was decreased by 1 semitone and the end-frequency was increased with 12 Hz. Thus, the seventh stimulus at the 'fear' endpoint of the continuum had an excursion size of 4 semitones and an end-frequency of 222 Hz. The duration of the utterances created in this way was then linearly compressed. The duration of the utterance at the 'happy' endpoint was left at 100% (i.e., 1.58 sec.) and the duration of each successive stimulus in the continuum was decreased with 2% so that duration at the 'fear' endpoint was 88% (i.e., 1.39 sec). All pitch and time manipulations were based on direct waveform manipulations (PSOLA, Charpentier & Moulines, 1989) so that the tokens sounded natural. These auditory stimuli were played directly from hard disk and presented at a comfortable listening level over loudspeakers.

Visual materials. The main visual stimuli consisted of two static facial expressions ('happy' and 'fearful') of the male actor who had spoken the original sentences. Each black-and-white picture was positioned in a 23 x 16-cm frame against a dark background. A face was presented at the onset of the voice and stayed on the screen for the duration of the utterance. Participants sat at a distance of approximately 60-cm from a PC screen on which the pictures were presented.

Design and Procedure. The same vocal and visual stimuli were used in three conditions. In the STANDARD condition, subjects heard one of the seven auditory

stimuli while a happy or a fearful picture of the face was shown. Participants pressed one of two keys to indicate whether the emotion in the voice was happy or fearful, and were instructed to ignore the face when making this judgement. In the ADD condition, the same auditory and visual stimuli were used as in the standard condition, but in addition two visual digits, randomly chosen between 1 and 9, were superimposed on the middle of the forehead of the face, one after the other for 600 ms each during presentation of the voice and face. The onset of the first digit was synchronized with the onset of the face. The digits were white (12 point font) on a dark background frame. After participants had judged the emotion in the voice, they entered the sum of the digits on a keyboard. We also included a CONTROL-DIGITS condition in which the same display was used as in the add-condition, but participants were not required to add the digits, and only judged whether the emotion in the voice was happy or fearful. This control condition was included to check whether the requirement of performing a dual-task, or merely the visual presence of the digits in the display, was responsible for any difference between the STANDARD and ADD condition.

Each condition was presented in a single block of 70 experimental trials (5 repetitions of the 2 face x 7 voice combinations, in random order) preceded by two warm-up trials. Participants received all three conditions in counterbalanced order. Gaze and any eye movements were checked during testing by the experimenter via a closed video circuit. Participants were tested individually in a sound shielded booth, and testing lasted about 20 min.

### Results and Discussion

Participants correctly added the two numbers for the add-condition on 90% of the trials (range: 75%-98%). The proportion of 'happy' responses in the voice task was determined for each condition and participant, and was then submitted to an Analysis of Variance (ANOVA). In the ADD condition, only trials in which the correct sum of the two digits was entered were considered to be valid (including the incorrectly added trials had virtually no effect on the ANOVAs). In the 3 (Task) x 2 (Face) x 7 (Voice) ANOVA on the proportion of happy responses, there was an effect of Voice,  $F(6,90) = 51.84$ ,  $p < .001$ , because the number of 'happy' responses increased as the voice changed from the fearful toward the happy end of the

continuum (Figure 1, upper panel). There was a significant effect of Face,  $F(1,15) = 10.76$ ,  $p < .005$ , because more 'happy' responses were given when a 'happy' face was seen than a 'fearful' face. The theoretically important interaction between Task and Face was marginally significant,  $F(2,30) = 3.27$ ,  $p = .052$ , while all other interactions were non-significant. The interaction between Task and Face indicated that the impact of the face was somewhat different for the three conditions. In order to assess the impact of the face on the voice directly, we subtracted the percentage of happy responses when a fearful face was shown from the percentage of happy responses when a happy face was shown across all levels of the voice continuum. The thus computed visual impact in the STANDARD condition was .15, in the CONTROL-DIGITS condition it was .14, while in the ADD condition it was .22. Separate  $t$ -test (two-tailed) showed that the visual impact in the ADD condition was somewhat bigger than in the STANDARD condition,  $t(16) = 2.01$ ,  $p = .06$ , and bigger than in the CONTROL-DIGITS condition,  $t(16) = 2.17$ ,  $p < .05$ . There was no significant difference between the STANDARD and CONTROL-DIGITS condition ( $p > .20$ ). Quite unexpectedly, there is thus a suggestion that the impact of the face was if anything bigger when there was an additional task. One interpretation of this finding is that while the cross-modal effect from the face on the voice arises automatically, subsequently ignoring the face requires attention. Participants may then find it harder to ignore the face when there is an additional task.

We explored this possibility further by changing the nature of the additional task. A potential difficulty with the ADD task is that, due to its overlearned nature, it may have been too easy so that resources were left. Another potential problem is that the additional task did not draw sufficiently on perceptual resources in particular. In Experiment 2, we therefore increased the perceptual demands of the additional task. Participants had to continuously scan a rapid serial visual presentation of digits for the occurrence of one or more zeroes, a task which is known to be very attention demanding (see e.g., Joseph, Chun, Nakayama, 1996). If the cross-modal effect on emotional judgements for the voice remains uninfluenced by the additional task, it would seem likely that the cross-modal effect itself does not require attentional resources.

## Experiment 2

The experiment was similar to the previous one, except that the additional task now required participants to detect (and count) the number of zeroes in a rapid serial visual presentation of digits.

### Method

Participants. A new group of 21 college students was tested. They received course credits for their participation.

Materials and Design. The STANDARD condition was as before, but a COUNT condition replaced the add-condition of Experiment 1. In this COUNT condition, a sequence of visual digits was now superimposed on the face. These digits were presented at a rate of 5 per s during the time that the face was shown. Within a sequence, no, one, or two zeroes appeared (at random serial positions) and participants were asked to count the number of zeroes within the sequence. They judged the emotion in the voice, and then pressed another key to indicate the number of zeroes. In the DIGITS control condition, the same display was shown as in the COUNT condition, but subjects could ignore the digits and only judged the voice.

### Results and Discussion

The data were analyzed as before. Participants correctly counted the number of zeroes in the COUNT condition on 90% of the trials (range: 73%-100%). A trial in the COUNT condition was only considered to be valid if the correct number of zeroes was entered. A 3 (Task) x 2 (Face) x 7 (Voice) ANOVA was performed on the proportion of happy responses with Task, Voice, and Face as within-subjects variables (Figure 1, middle panel). As expected, the number of 'happy' responses increased as the voice changed from the fearful towards the happy end of the continuum,  $F(6,120) = 83.76$ ,  $p < .001$ . There was a main effect of Face,  $F(1,20) = 26.41$ ,  $p < .001$ , because more 'happy' responses were given when a 'happy' face was shown rather than a 'fearful' face. There was also a significant interaction between Face and Voice because the effect of the face was bigger in the ambiguous regions of the voice continuum,  $F(6,120) = 5.24$ ,  $p < .001$ . So far these data replicate the results of de Gelder and Vroomen (2000). Most importantly, there was no main effect of task, and all other interactions were non-significant. The visual impact of the



face on voice judgements was .05 in the STANDARD condition, .08 in the DIGITS control condition, and .07 in the COUNT condition. Separate  $t$ -test comparing these visual impacts among each other were all non-significant (all  $p$ 's > .10).

Thus, there was again no sign that the additional task decreased the cross-modal impact of the face on the voice, suggesting that the cross-modal effect does not require attentional resources in this sense. However, so far we used additional tasks that involved visual displays. One might argue that these displays could themselves draw attention to the visual modality in general, thereby increasing the influence of the visual face on judgement of heard emotion. To assess this possibility further, we used an additional auditory task instead for Experiment 3.

### Experiment 3

This experiment was similar to the previous ones, except that participants now had an additional auditory task in which they had to discriminate a 500 Hz from a 540 Hz tone that was presented together with the voice.

#### Method

Participants. A new group of 10 college students was tested. They received course credits for their participation.

Materials and Design. The STANDARD condition was as in the previous experiments. In the PITCH condition, there was a 500 Hz (low) or 540 Hz (high) tone (100 ms in duration) that was added to the original waveforms either at 350 or 950 ms after sentence onset. Participants not only had to judge the emotion in the voice, they also had to decide whether they had heard a low or high tone by pressing one of two designated keys. In the SOUND condition, they heard the same stimuli as in the PITCH condition, but participants could ignore the tones. All other details were as before.

#### Results and Discussion

The additional task was more difficult than in the previous experiments, with participants correctly judging the pitch of the tone in the PITCH condition on only 70% of the trials (range: 54%-99%; in Experiments 1 and 2, the average performance on the additional task had been 90% and 91%, respectively). As in the

previous experiments, trials of the PITCH condition were only considered to be valid if the pitch of the tone was correctly judged. However, this now amounted to discarding a larger number of trials as before. We therefore checked that this did not influence our results by performing another analysis in which incorrectly judged trials were not discarded. The results of both analyses were almost identical, and we therefore report only the former.

In a 3 (Task) x 2 (Face) x 7 (Voice) ANOVA on the proportion of happy responses, there was an effect of Voice,  $F(6,54) = 120.72$ ,  $p < .001$ , because the number of 'happy' responses increased as the voice changed from the fearful toward the happy end of the continuum (Figure 1, lower panel). There was a main effect of Face,  $F(1,9) = 5.68$ ,  $p < .05$ , because more 'happy' responses were given when a 'happy' face was seen than a 'fearful' face. Importantly, all other terms were non-significant (all  $p$ 's  $> .10$ ). The visual impact in the STANDARD condition was .10, in the SOUND condition it was .10, and in the PITCH condition it was .08. Separate  $t$ -test showed that none of these effects were significantly different from each other (all  $p$ 's  $> .10$ ). The cross-modal effect was thus again independent of whether or not a difficult secondary task had to be performed (now in audition) in addition to the emotion judgement.

### General Discussion

In the present study, we examined whether the integration of visual and auditory information about emotions requires attentional resources. Participants judged whether a voice expressed happiness or fear, while being required to ignore a static face expressing either one of these emotions. As observed previously, the visual information had an impact on judgments of the heard emotion despite instructions to ignore the face (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). Since manipulating instructions, as in previous studies, is only an indirect and imperfect way of understanding the role of attention, we set up three experiments that targeted specifically the role of attentional load in cross-modal emotion perception. Here we observed that the cross-modal influence from a face on the rating of affect for a voice was obtained independent of whether or not participants performed a demanding additional task, and independent of the nature of this additional task (i.e., adding two numbers, processing a rapid serial visual

presentation of digits, or judging the pitch of a tone). This suggests that the integration of visual and auditory information about emotions may be a mandatory process, unconstrained by attentional resource.

From an ecological point of view, automatic integration of affective inputs across sensory modalities makes adaptive sense, given the biological significance of emotional stimuli, plus the fact that combining different sources of information (as in the face and voice) should usually lead to more accurate judgements and more appropriate behavior. However, one should be cautious interpreting a cross-modal bias as a true cross-modal integration effect. An alternative possibility is that the visual and auditory information were not integrated, but were processed in parallel and independently of each other. It may be that on some trials participants rely on the non-target source of information, for instance when they are less sure about the information in the target modality. Such a 'response bias' (rather than a changed perception) might predict the same psychometric pattern as was obtained here, and so the present data cannot rule out such an account. Yet there is some literature showing that response bias on its own cannot account for all the data on cross-modal emotion perception.

This includes evidence that facial expressions are processed rapidly, even when conditions of normal vision do not obtain. This arises when stimulus awareness can be ruled out, as a consequence of visual masking of the face stimulus (Morris, Oehman & Dolan, 1998; Whalen et. al., 1998). More dramatically, patients with unilateral striate cortex lesions (producing hemianopia, but with residual unconscious vision or 'blindsight') can still process facial expressions when the face is presented in the consciously blind field (de Gelder, Pourtois, van Raamsdonk, Vroomen, Weiskrantz, 2001; de Gelder, Vroomen, Pourtois, Weiskrantz, 1999; Morris, de Gelder, Weiskrantz & Dolan, 2001). Similarly, patients with unilateral parietal damage suffering from hemineglect process facial expressions when the face is presented in the contra-lesional field (Driver & Vuilleumier, in press).

There is also evidence that faces can exert relatively early cross-modal effects when combined with voices. Recently, we studied the time course of the crossmodal bias effect by recording modifications of auditory Evoked Potentials, due to the presence of a facial expression when subjects listened to emotional voices (Pourtois, de Gelder, Vroomen, Rossion, & Crommelink, 2000). The latencies corresponding to

the visual bias indicate an effect that may be too early in time to depend on attentional resources or response biases (i.e., 110 ms post-stimulus onset). An initial event-related functional MRI exploration on the neuroanatomical basis of the crossmodal effect suggests that emotion-dependent activation of the amygdala to a fearful face can be increased by the presence of a congruent fearful emotion in the voice (as compared to pairing with a happy voice; Dolan, Morris & de Gelder, 2001). "The emerging picture is that integration of emotional information provided by the voice and the face may be integrated directly via a thalamic-amygdala route, while routes from amygdala to cortex may explain any subsequent attentional modulation (Ledoux, 1996).

It remains to be investigated whether the presumed automaticity of the cross-modal effect is a unique feature of face (or emotion) processing, or whether it extends to other domains as well. A partial answer to this may come from research on cross-modal interactions in the spatial domain, or the so-called ventriloquist-effect. The ventriloquist-effect can be observed when participants are instructed to localize sounds while ignoring spatially discordant lights. Localisation is usually shifted in the direction of the light, despite instructions to focus only on the target sound (e.g., Bertelson & Radeau, 1981). The ventriloquist effect is independent of whether the participant focuses attention on the distracter light or not (Bertelson, Vroomen, de Gelder, & Driver, 2000) and independent from where automatic visual attention, as captured by a unique element in a visual display, is directed (Vroomen, Bertelson, & de Gelder, 2001). It is also obtained in patients suffering from hemispatial neglect who are unaware of the presence of a stimulus in the unattended field (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000). It thus seems that at least some interactions at the core of cross-modal effects (in both affective realms, as here, and also spatial realms, as for ventriloquism) can take place at a stage of processing before attention comes into play (see also Vroomen & de Gelder, 2000, Driver, 1996).

## References

- Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*, 1634-1642.
- Bertelson, P. & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*, 578-587.
- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321-332.

- Charpentier, F., & Moulines, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis. *Proceedings of Eurospeech*, *89*, 2, 13-19.
- de Gelder, B., Pourtois, G., van Raamsdonk, M., Vroomen, J. & Weiskrantz, L. (2001). Unseen stimuli modulate conscious visual experience: Evidence from inter-hemispheric summation. *NeuroReport*, *12*, 385-391.
- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*, 289-311.
- de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *NeuroReport*, *10*, 3759-3763.
- de Gelder, B., Vroomen, J., & Teunisse, J-P. (1995). Hearing smiles and seeing cries: The bimodal perception of emotions. Paper presented at the 36<sup>th</sup> annual meeting of the Psychonomic Society, Los Angeles, CA.
- Dolan, R., Morris, J., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences*, *98*, 10006-10010.
- Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lipreading. *Nature*, *381*, 66-68.
- Driver, J., and Vuilleumier, P. (in press). Unconscious processing in neglect and extinction. In B. de Gelder, E. de Haan & C. A. Heywood (Eds.), *Out of mind*. Oxford University Press.
- Joseph, J. S., Chun, M. M., & Nakayama, K. (1996). Attentional requirements in a "preattentive" feature search task. *Nature*, *379*, 805-807.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451-468.
- Lavie, N. (2000). Selective attention and cognitive control. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention & Performance XVIII* (pp. 175-194). Cambridge, MA: MIT Press.
- Ledoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY: Simon and Schuster.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge: MA: MIT Press.
- Massaro, D. W. & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*, 215-221.
- Morris, J. S., de Gelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124*, 1241-1251.
- Morris, J. S., Oehman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Science*, *96*, 1680-1685.
- Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelink, M. (2000). The time course of intermodal binding between seeing and hearing affective information. *NeuroReport*, *11*, 1329-1333.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-190.
- Vroomen, J., Bertelson, P., de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*, 651-659.
- Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583-1590.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*, 411-418.
- Wickens, D. D. (1984). Processing resources in attention. In R. Pararsuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). Orlando, FL: Academic Press.

## Figure captions

Figure 1. The proportion of 'happy' responses as a function of the auditory continuum, separately for each of the additional tasks. The three upper panels show the data for Experiment 1, the three middle panels for Experiment 2, and the three lower panels for Experiment 3.

