

Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures

Beatrice de Gelder^{*†‡}, Gilles Pourtois^{*†}, and Lawrence Weiskrantz[§]

^{*}Donders Laboratory for Cognitive and Affective Neuroscience, University of Tilburg, NL-5000 LE Tilburg, The Netherlands; [†]Laboratory of Neurophysiology, Louvain University, 1200 Brussels, Belgium; and [§]Department of Experimental Psychology, Oxford University, Oxford OX1 2JD, United Kingdom

Contributed by Lawrence Weiskrantz, January 11, 2002

Multisensory integration is a powerful mechanism for increasing adaptive responses, as illustrated by binding of fear expressed in a face with fear present in a voice. To understand the role of awareness in intersensory integration of affective information we studied multisensory integration under conditions of conscious and nonconscious processing of the visual component of an audiovisual stimulus pair. Auditory-event-related potentials were measured in two patients (GY and DB) who were unable to perceive visual stimuli consciously because of striate cortex damage. To explore the role of conscious vision of audiovisual pairing, we also compared audiovisual integration in either naturalistic pairings (a facial expression paired with an emotional voice) or semantic pairings (an emotional picture paired with the same voice). We studied the hypothesis that semantic pairings, unlike naturalistic pairings, might require mediation by intact visual cortex and possibly by feedback to primary cortex from higher cognitive processes. Our results indicate that presenting incongruent visual affective information together with the voice translates as an amplitude decrease of auditory-event-related potentials. This effect obtains for both naturalistic and semantic pairings in the intact field, but is restricted to the naturalistic pairings in the blind field.

Observers may respond to the expression of a face even though they are unable to report seeing it. For example, when confronted with a backward-masked, hence unseen, angry face, observers nevertheless give a reliable skin-conductance response (1). Facial expressions of fear increase the activation level of the amygdala, and this effect is lateralized as a function of whether or not the observer can perceive the faces consciously with right amygdala activation reported for seen, and left amygdala for unseen, masked presentation (2). Faces displaying emotional expressions followed by a backward mask lead to increased activities among the left amygdala, pulvinar, and superior colliculus (3). Because these structures are still intact in case of brain damage restricted to striate cortex, we conjectured that these patients may also recognize affective stimuli just as they can recognize some elementary visual stimulus attributes in the absence of awareness, a phenomenon called blindsight. Accordingly, we have shown that patient GY, who has a lesion in his striate cortex, was able to discriminate between facial expressions he could not see and was not aware of (affective blindsight; refs. 4 and 5). This result is consistent with the notion that, in normal subjects, unseen facial expressions are processed by means of a subcortical pathway involving right amygdala, pulvinar, and colliculus, whereas processing of seen faces increases connectivity in fusiform and orbitofrontal cortices (3). Subsequently we confirmed the importance of the noncortical route in relation to the difference between seen/aware vs. unseen/unaware facial fear expression and hemispheric side of activation (6). The extension of affective blindsight to stimuli other than faces has not, to our knowledge, previously been envisaged.

A different role of the amygdala concerns its function in cross-modal binding. Animal studies in which the crossmodal function of the amygdala has been demonstrated used reward-based condi-

tioning to establish crossmodal pairing (7, 8). We recently explored pairing between affective expressions of the face and the voice in humans (9, 10) by measuring the crossmodal bias exercised by the face on recognition of fear in the voice and vice versa. The pattern of brain activations suggests that the amygdala plays a critical role in this process of binding affective information from the voice and the face. This was indicated by increased activation to fearful faces when they were accompanied by voices that expressed fear (11). Electrophysiological recordings of the time course of the cross-modal bias from the face to the voice indicated that such intersensory integration takes place on-line during auditory perception of emotion (12, 13) and mainly translates as amplitude change of exogenous auditory components [i.e., early auditory potential (N1) and mismatch negativity (MMN)]. Importantly, facial expressions can bias perception of an affective tone of voice even when the face is not attended to (14) or cannot be perceived consciously, as is the case with brain damage to occipitotemporal areas (15). These latter findings indicate that the crossmodal effect does not depend on conscious recognition of the visual stimuli. Yet residual vision in the sense of covert processing (found in some cases of prosopagnosia with typical occipitotemporal damage) is quite different from a neuroanatomical point of view from loss of stimulus awareness caused by striate cortex lesion accompanied by blindsight. Unlike more anterior areas, striate cortex may play a critical role in conscious perception (16) and/or, because of its involvement in feedback projections from higher visual areas (17), a role that may be critical for audiovisual binding.

Here we asked whether unseen visual stimuli could still cross-modally influence auditory processing in the absence of normal striate cortex function. The results would provide a crossmodal approach to existing indirect methods for testing covert processes that have so far been restricted to single channel methods (18). This is the first issue we addressed. Our second question concerns the perception of other emotion-inducing objects besides faces. Visual stimuli with clear affective valence, such as pictures of spiders, snakes, or food, have frequently been used to study processes involved in fear perception (19, 20); but as of yet there is no evidence that other fear-inducing stimuli besides faces can be processed in the case of striate cortex lesion. If this is indeed the case, an important question is whether nonconscious perception of emotional pictures can influence recognition of emotional voices similarly to what we predict to be the case for nonconsciously perceived faces. Alternatively, the processes underlying an affective scene-voice pairing may require higher-order mediation because such a pairing is based on semantic properties they might share. Blindsight patients are of crucial importance for addressing this issue because they create conditions for testing of audiovisual

Abbreviations: EEG, electroencephalogram; EOG, electro-oculographic.

[†]To whom reprint requests should be addressed. E-mail: b.degelder@kub.nl.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

pairings based on semantic associations when the subjects are either aware or unaware of the visual stimuli.

We studied affective blindsight for emotional pictures in two patients (GY and DB) with unilateral striate lesions. We recorded electrical brain responses to presentation of audiovisual stimulus pairs to the intact and the blind visual field while the subject was attending to the auditory part of the stimulus pair and making a task-irrelevant judgment of the gender of the voice (21). We had two *a priori* hypotheses. First, we predicted a decrease in amplitude of N1 for the naturalistic audiovisual pairings (voice–face) for the condition in which the emotionally congruent face was replaced by an incongruent one. Second, we expected that a similar effect would also be found for the semantic pairings (voice–scene) when the visual stimulus was presented to the intact hemisphere and was fully processed and consciously perceived. But we did not expect that an emotional picture presented to the blind field would influence recognition of voice expressions even if patients had blindsight for the emotional valence. It was presumed that the subcortical and amygdala-based activation postulated to be responsible for binding of emotional faces and voices would not be sufficient for crossmodal binding of voice–scene pairings. The latter would require mediation from extra-amygdala representations of the fear-inducing stimuli that cannot be accessed with striate cortex damage.

Methods

Experimental Sequence. The electroencephalogram (EEG) experiment took place first and was followed some months later by the behavioral experiment. During the behavioral experiment, eye movements were monitored by closed-circuit TV. During the EEG experiment, eye movements were monitored by means of electrodes attached to the orbits of the eyes calibrated to record deviations from fixation.

Subjects. Two patients with blindsight (GY and DB) were studied. Patient GY is a 45-year-old male who sustained damage to the posterior left hemisphere of his brain by head injury (a road accident) when he was 7 years old. The lesion (see ref. 22 for an extensive structural and functional description of the lesion) invades the left striate cortex (i.e., medial aspect of the left occipital lobe, slightly anterior to the spared occipital pole, extending dorsally to the cuneus and ventrally to the lingual, but not the fusiform gyrus) and surrounding extrastriate cortex (inferior parietal lobule). The location of the lesion is functionally confirmed by perimetry field tests (see ref. 23 for a representation of GY's perimetric field and ref. 24 for a comparison).

DB is a 61-year-old male from whom an arterious venous malformation in his medial right occipital lobe was surgically removed when he was 33. As clinical visual symptoms were first noted in his teens, it is presumed that the nonmalignant tumor had been present for several years, perhaps even prenatally. The excision extended ≈ 6 cm anterior to the occipital pole and included the major portion of the calcarine cortex on the medial surface. The operation produced a homonymous macula-splitting hemianopia, with a crescent of preserved vision at the periphery of the upper quadrant. Because metal clips were used in the surgical procedure (including an aneurysm clip), MRI scans are not possible. Some information can be discerned from a computed tomography (CT) scan, which is clear in demonstrating no remaining striate cortex in the upper bank of the calcarine fissure, corresponding to the lower quadrant of the impaired hemifield. It is presumed, from surgical notes, that the lower bank was also destroyed, although the CT is severely distorted. He was studied in a series of psychophysical tests in 1974 by Weiskrantz and collaborators (25), leading to the original characterization of “blindsight,” and then subsequently for more than 10 years, with results summarized in 1986 by Weiskrantz (26). Contact was broken for some years, but he has been seen again in a series of ongoing studies since 1999. His blindsight capacities are essentially the same as originally described.

Behavioral Experiment. GY and DB were presented with visual stimuli in a direct guessing paradigm used similarly for facial expressions and emotional pictures. Visual materials consisted of black and white photographs of facial expressions and emotional pictures. The face set consisted of 12 images (six individuals, once with a happy facial expression and once with a fearful facial expression; ref. 27). Emotional pictures were 12 black and white static pictures selected from the International Affective Picture System (20) because they presented a homogenous class of nonfacial stimuli. There were 6 negative pictures (snake, pit bull, 2 spiders, roaches, and shark; mean valence: 4.04 ± 0.4) and 6 positive ones (porpoises, bunnies, lion, puppies, kitten, and baby seal; mean valence: 7.72 ± 0.45). Stimulus properties (luminance and mean size), presentation modalities, and task requirements for the two types of visual materials were designed so as to be maximally comparable. Visual stimuli were presented on a 17-inch screen. Mean size of the face pictures was 6-cm width by 8-cm height (sustaining a visual angle of 5.73° horizontally by 7.63° vertically) and 8-cm width by 6-cm height for the emotional pictures (sustaining a visual angle of 7.63° horizontally by 5.73° vertically). Mean luminance was 25 cd/m^2 and less than 1 cd/m^2 for the background and ambient light.

A trial consisted of the presentation of the visual stimulus (face or scene) in the blind field lasting for 1,250 ms. The subject was instructed to make a two-alternative forced-choice response between happy and fearful and to respond with their dominant hand by pressing the corresponding button of a response box. He was instructed to make his judgement before the end of the stimulus presentation (within 1,250 ms after stimulus onset) to eliminate any interference by offset transients (in DB, the offset of a stimulus under certain conditions gives rise to after-images; ref. 28) and to maintain fixation during the block. Intertrial interval was 1,000 ms. Two blocks of 48 trials (24 faces and 24 pictures) were randomly presented (four blocks with GY).

EEG Experiment. Stimuli consisted of pairings of an auditory and a visual component presented simultaneously. Visual materials were identical those of the behavioral experiment. Auditory materials consisted of 12 bisyllabic spoken words obtained with the following procedure. Six male and six female actors were instructed to pronounce a neutral sentence (“they are traveling by plane”) in an emotional tone of voice (either happy or fearful). Speech samples were recorded on a Digital Audio Tape recorder and subsequently digitized and amplified (using SOUNDEDIT 16 1.0 B4 running on Macintosh). Tokens of the final word “plane” were then selected by using SOUNDEDIT, resulting in a total of 24 samples (12 actors \times 2 tones of voice). They were presented in a pilot study to 8 volunteers (4 males and 4 females), none of whom participated in the experiment. Subjects labeled each fragment as “happy,” “fearful,” or “don't know.” Based on their recognition rates, 12 fragments were selected (mean recognition rate 74% correct). Audiovisual pairings were obtained by combining a sound fragment with either a facial expression or an emotional scene. Pairings were either congruent or incongruent. There were 12 pairs for each visual condition. Trials were generated by using the STIM software running on a PC Pentium II. Visual stimuli were presented as described above. Sounds were delivered over two loudspeakers placed on each side of the screen at a mean sound level of 72 dB.

A trial consisted of the presentation of the visual stimulus followed after 900 ms by the voice fragment (duration 381 ± 50 ms), during which the image remained on the screen (Fig. 1). This delay between visual and auditory stimulus onset was introduced to reduce interference from the brain response elicited by the visual stimuli (12, 13). The 12 fragments used for the congruent and incongruent conditions as well as for the two visual conditions were always identical (the only difference was the emotional valence of the face/emotional picture they were paired with). Four blocks of 192 trials (48 stimuli consisting of 24 face–voice pairs and 24

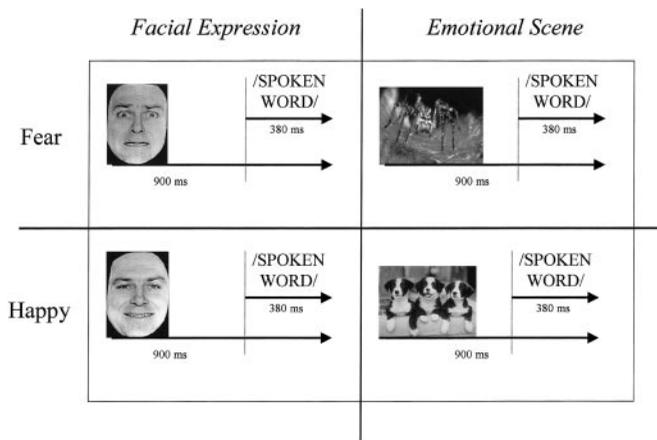


Fig. 1. Experimental design (EEG experiment). Subjects viewed either happy or fearful visual stimulus (a face or a scene) while listening to the word “plane” spoken in either happy or fearful tones. There were eight resulting conditions: congruent happy face, congruent happy scene, incongruent happy face, incongruent happy scene, congruent fearful face, congruent fearful scene, incongruent fearful face, and incongruent fearful scene. Subjects were instructed to identify the gender of the voice as either male or female (by means of a button press).

scene–voice pairs, half congruent and half incongruent, repeated four times) were randomly presented in each hemifield. Intertrial interval was 1,000 ms. Subjects were tested in a dimly lit room seated 60 cm away from the screen. The subjects were instructed to fixate a central cross on the screen and not to pay attention to the visual stimuli appearing in the periphery. Their task was to perform a gender decision on the voices.

Data Acquisition. Auditory event-related brain potentials were recorded and processed by using a 64-channel acquisition system (Neuroscan). Horizontal and vertical electro-oculographic (EOG) monitoring was realized by using four facial bipolar electrodes placed on the outer canthi of the eyes and in the inferior and superior areas of the orbit. Scalp EEG was recorded from 58 electrodes mounted in an Electrocap (10–20 system) with a linked-mastoids reference, amplified with a gain of 30,000 and bandpass-filtered at 0.01–100 Hz. Impedance was kept below 5 k Ω . EEG and EOG were continuously acquired at a rate of 500 Hz.

Data Analysis. After removal of EEG and EOG artefacts (epochs with EEG or EOG exceeding $\pm 70 \mu\text{V}$ were excluded from the averaging), epoching was made 100 ms before auditory stimulus onset and for 924 ms after stimulus presentation. Data were low-pass filtered at 30 Hz. Maximum amplitudes of auditory event-related brain potentials (AEPs) were measured relative to a 100-ms prestimulus baseline and assessed by using repeated-measures ANOVAs and Student’s *t* tests. To carry out these ANOVAs, the four blocks for each hemifield for each patient were considered as nonrepeated and entered in the analyses as independent measures. Statistical analyses were focused on the amplitude modulation of the central negative deflection occurring 110 ms after auditory stimulus and called the N1 component. For each condition, the mean amplitude of the N1 component was measured relative to the maximum negativity at nine adjacent central electrodes (C3A, CzA, C4A, C3, Cz, C4, C3P, PzA, and C4P) in the 90- to 150-ms interval. These nine electrodes with a central topography were chosen because they are best suited to record early exogenous AEPs simultaneously at several scalp positions.

Results

Behavioral Results. For 40 of 96 trials, DB responded after the stimulus offset (despite instructions to respond during the stimulus

presentation) leaving only 56 trials for the analysis. He was significantly above chance in forced choice guessing the affective content of the visual stimuli presented in his blind visual field and equally well for the two visual categories. For emotional pictures, DB made 5/22 errors or 77% correct [$\chi^2_{(1)} = 15.7, P < 0.001$]. For facial expressions, he made 7/34 errors (4 happy and 3 fear) or 79% correct [$\chi^2_{(1)} = 18.4, P < 0.001$].

GY was at chance level for facial expressions (45/94 errors or 52% correct), but was above chance level for emotional pictures [34/94 errors or 64% correct, $\chi^2_{(1)} = 7.84, P < 0.01$] with 12/48 errors (75% correct) for fear and 22/46 (52%) for happy trials.

EEG Results. Behavioral data collected online indicate that as expected the task irrelevant gender decision was easy (95% for GY and 85% for DB). Our main interest focused on amplitude changes of the auditory N1 component that would be a function of the type of visual stimulus. Significant main effects of laterality, anteriority, or emotion, as well as interactions between these variables, are therefore reported only when they significantly interact with one of the main experimental variables (hemifield, visual category, and congruency).

Intact visual field. Results supported by statistical analyses revealed that for GY incongruent audiovisual pairs elicited a lower N1 component than congruent pairs at several electrode positions for both types of visual stimuli. The ANOVA performed on the mean amplitude of the N1 component at nine electrode positions with the factors *Visual Category* (nonfacial context vs. facial context), *Emotion* (happy vs. fearful), *Congruency* (congruent vs. incongruent pairs), *Anteriority* (anterior, central, and posterior), and *Laterality* (left, mid-line, or right) indicated a significant interaction visual category \times anteriority [$F(2,6) = 5.3, P < 0.05$] and a significant interaction between the five variables [$F(4,12) = 5.7, P < 0.01$]. For scene–voice pairs, the 2 (emotion) \times 2 (congruency) \times 3 (anteriority) \times 3 (laterality) interaction approaches significance [$F(4,12) = 2.6, P = 0.09$]. Post hoc *t* tests for the two emotions averaged together showed that for four of nine electrode positions [C3A, $t(3) = 2.12, P = 0.056$; C3, $t(3) = 2.2, P = 0.049$; C3P, $t(3) = 4.1, P < 0.005$; C4P, $t(3) = 2.14, P = 0.054$] incongruent pairs elicit a lower N1 component than congruent pairs (Table 1 and Fig. 2a). For face–voice pairs, the analysis revealed a significant interaction emotion \times congruency \times anteriority [$F(2,6) = 6.71, P < 0.05$] and a significant interaction congruency \times anteriority \times laterality [$F(4,12) = 3.43, P < 0.05$]. For 5 of 9 electrode positions [C3A, $t(3) = 3.1, P < 0.01$; Cz, $t(3) = 3.02, P < 0.05$; C4, $t(3) = 3.04, P < 0.05$; PzA, $t(3) = 5.18, P < 0.001$; C4P, $t(3) = 3.03, P < 0.05$] post hoc *t* tests showed that incongruent pairs elicited a lower N1 component than congruent pairs (Table 1 and Fig. 2a).

Results for DB were very similar. For scene–voice pairs, post hoc *t* tests showed that for two of nine electrode positions [PzA, $t(3) = 2.13, P = 0.055$; C4P, $t(3) = 4.53, P < 0.001$], incongruent pairs elicit a lower N1 component than congruent pairs (Table 2 and Fig. 2b). For face–voice pairs, post hoc *t* tests showed that for three of nine electrode positions [C3, $t(3) = 2.17, P = 0.051$; PzA, $t(3) = 3.43, P = 0.005$; C4P, $t(3) = 2.92, P < 0.05$], incongruent pairs elicit a lower N1 component than congruent pairs (Table 2 and Fig. 2b).

Blind visual field. For GY, the ANOVA on the mean amplitude of the N1 component for visual presentations in the blind visual field revealed a significant interaction visual category \times emotion \times anteriority [$F(2,6) = 5.58, P < 0.05$]. For scene–voice presentations, the ANOVA did not disclose any significant interaction with congruency. For face–voice presentations, the interaction congruency \times anteriority approached significance [$F(2,6) = 3.94, P = 0.081$] indicating that incongruent face–voice pairs elicited a lower N1 component at central [$t(3) = 2.39, P = 0.054$] and centroparietal leads [$t(3) = 3.93, P < 0.01$] but not at anterior leads [$t(3) < 1$] than congruent face–voice pairs (Table 1 and Fig. 2a). The decrease in amplitude of the N1 component for incongruent face–voice pairs is

Table 1. Patient GY: Mean amplitudes of N1 components

Hemifield	Electrode position	Condition	Peak amplitude, μV	t (df = 3)
LVF (intact VF)	P3P	Face congruent	-12.1	
		Face incongruent	-11.92	0.51
		Scene congruent	-12.25	
	PzP	Scene incongruent	-10.59	4.09**
		Face congruent	-9.22	
		Face incongruent	-7.38	5.18**
	P4P	Scene congruent	-9.34	
		Scene incongruent	-9.03	0.77
		Face congruent	-7.14	
RVF (blind VF)	P3P	Face incongruent	-6.06	3.03**
		Scene congruent	-8.13	
		Scene incongruent	-7.26	2.14*
	PzP	Face congruent	-12.1	
		Face incongruent	-11.54	1.22
		Scene congruent	-12.36	
RVF (blind VF)	PzP	Scene incongruent	-12.44	0.21
		Face congruent	-9.4	
		Face incongruent	-8.77	1.41
	P4P	Scene congruent	-8.57	
		Scene incongruent	-8.85	0.75
		Face congruent	-8.55	
P4P	Face incongruent	-7.03	3.35**	
	Scene congruent	-8.18		
	Scene incongruent	-8.22	0.12	

Mean amplitude of the N1 component for three electrode positions in each visual condition (scene or face, congruent or incongruent) and for each visual hemifield (left/intact vs. right/blind). *, Indicates a P value for the t test < 0.05 ; **, $P < 0.01$.

Table 2. Patient DB: Mean amplitudes of N1 components

Hemifield	Electrode position	Condition	Peak amplitude, μV	t (df = 3)
LVF (blind VF)	P3P	Face congruent	-8.63	
		Face incongruent	-8.43	0.65
		Scene congruent	-6.73	
	PzP	Scene incongruent	-6.84	0.24
		Face congruent	-8.59	
		Face incongruent	-8.54	0.15
	P4P	Scene congruent	-6.89	
		Scene incongruent	-6.65	0.55
		Face congruent	-8.39	
RVF (intact VF)	P3P	Face incongruent	-7.49	2.93**
		Scene congruent	-6.41	
		Scene incongruent	-5.97	0.99
	PzP	Face congruent	-7.74	
		Face incongruent	-7.63	0.27
		Scene congruent	-6.89	
RVF (intact VF)	PzP	Scene incongruent	-7.17	0.96
		Face congruent	-8.1	
		Face incongruent	-6.62	3.43**
	P4P	Scene congruent	-7.12	
		Scene incongruent	-6.49	2.13*
		Face congruent	-7.88	
P4P	Face incongruent	-6.35	3.55**	
	Scene congruent	-7.8		
	Scene incongruent	-6.48	4.53**	

Mean amplitude of the N1 component for three electrode positions in each visual condition (scene or face, congruent or incongruent) and for each visual hemifield (left/blind vs. right/intact). *, Indicates a P value for the t test < 0.05 ; **, $P < 0.01$.

maximum at the right centroparietal leads C4P [$t(3) = 3.35, P < 0.01$].

For DB, the ANOVA disclosed a significant interaction visual category \times anteriority [$F(2,6) = 9.01, P < 0.05$], a significant interaction visual category \times emotion \times anteriority [$F(2,6) = 23.41, P < 0.005$] and a significant interaction visual category \times congruency \times anteriority \times laterality [$F(4,12) = 3.47, P < 0.05$]. For scene-voice pairs, none of the nine post hoc t tests is significant

(Table 2 and Fig. 2b). For face-voice pairs, post hoc t tests showed that for three (all situated in the right hemisphere) of nine electrode positions [C4A, $t(3) = 3.48, P < 0.005$; C4, $t(3) = 4.3, P < 0.005$; C4P, $t(3) = 2.93, P < 0.05$], incongruent pairs elicit a lower N1 component than do congruent pairs (Table 2 and Fig. 2b).

Comparison between intact and blind visual field. N1 amplitude observed in GY is highest at centroparietal leads (C3P, PzA, and C4P), and the incongruency effect generated by the visual stimulus

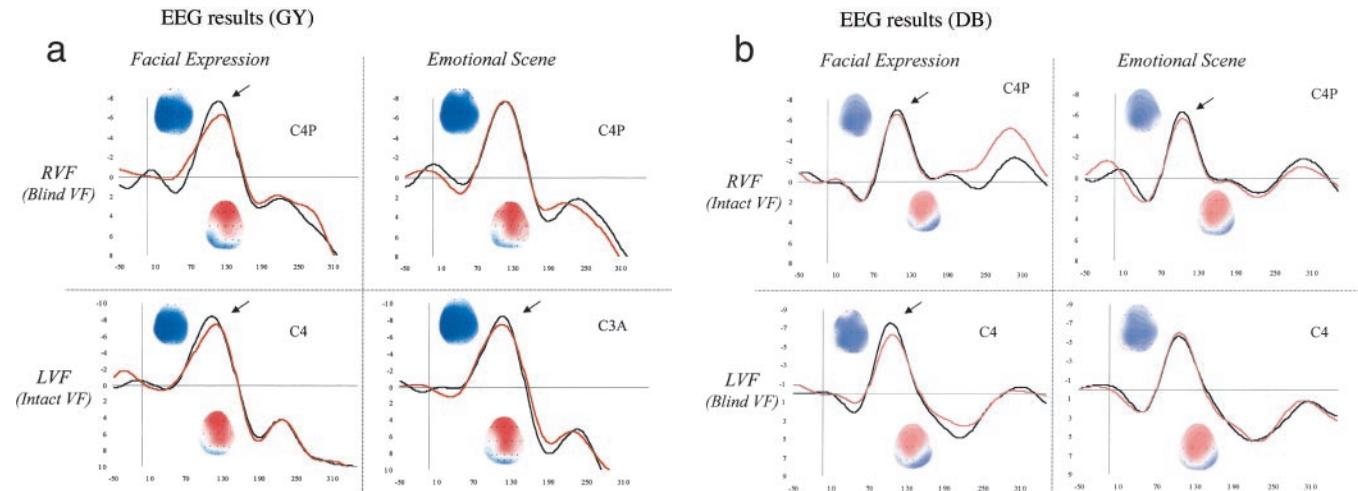


Fig. 2. The x axis represents the time in ms (from -50 ms before auditory stimulus onset to 310 ms after stimulus onset). The y axis represents the amplitude in μV from $-8 \mu\text{V}$ to $+8 \mu\text{V}$. Grand averaged auditory waveforms and corresponding topographies (horizontal axis) at central electrodes in each visual condition (congruent pairs in black, incongruent pairs in red) and for each visual hemifield are shown. For each topographical map (N1 and P2 components), the time interval is 20 ms and the amplitude scale goes from $-6 \mu\text{V}$ (in blue) to $+6 \mu\text{V}$ (in red).

(either a facial expression or an emotional picture) to auditory processing (as indexed by an amplitude modulation of the N1 component) appears to be stronger in the right hemisphere (C4P) than the left hemisphere (C3P) or the mid-line position (PzA) as revealed in the previous ANOVAs by significant interactions between anteriority and laterality.

To compare directly the effect of visual category on the early auditory processing in the intact and blind hemifield, a repeated-measures ANOVA was carried out on the mean amplitude of the N1 component at three electrode positions (C3P, PzA, and C4P) with the factors *Hemifield* (intact vs. blind), *Visual Category* (non-facial vs. facial context), *Emotion* (happy vs. fearful), *Congruency* (congruent vs. incongruent pairs), and *Laterality* (left, mid-line, or right). The analysis revealed a significant interaction visual category \times emotion [$F(1,3) = 9.37, P = 0.055$], a significant interaction hemifield \times visual category \times laterality [$F(2,6) = 15.74, P < 0.005$] and a significant interaction between the four variables [$F(2,6) = 7.74, P < 0.05$]. Importantly, in the intact visual field, incongruent pairs elicited a lower N1 component than congruent pairs whatever the category of the visual stimulus [at C4P, if the visual stimulus is a face, $t(3) = 2.79, P < 0.05$; if the visual stimulus is a scene, $t(3) = 2.26, P = 0.065$], whereas in the blind visual field, incongruent pairs elicit a lower N1 component than congruent pairs only if the visual stimulus is a face [$t(3) = 3.93, P < 0.01$] and not if the visual stimulus is a scene [$t(3) = 0.11, P > 0.9$].

Similar to that of patient GY, the amplitude of the N1 in DB is maximum at centroparietal leads (C3P, PzA, and C4P) and the incongruency effect from the visual stimulus to the early auditory processing seems to be stronger in the right hemisphere (C4P) than the left hemisphere (C3P) or the mid-line position (PzA) as revealed in the previous ANOVAs by significant interactions between anteriority and laterality. The repeated-measures ANOVA analysis revealed a significant interaction congruency \times laterality [$F(2,6) = 5.69, P < 0.05$] and a significant interaction visual category \times emotion \times congruency [$F(1,3) = 10.67, P < 0.05$]. Importantly, in the intact visual field, incongruent pairs elicited a lower N1 component than congruent pairs whatever the category of the visual stimulus [at C4P, if the visual stimulus is a face, $t(3) = 2.6, P < 0.05$; if the visual stimulus is a scene, $t(3) = 2.25, P = 0.065$], whereas in the blind visual field, incongruent pairs elicit a lower N1 component than congruent pairs only if the visual stimulus is a face [$t(3) = 2.93, P < 0.05$] and not if the visual stimulus is a scene [$t(3) = 0.75, P > 0.5$].

Discussion

The first noteworthy result is that both patients are able to nonconsciously discriminate the expression of emotional pictures. DB performed equally well in guessing the emotional attribute of faces and pictures. Results with GY are consistent with our previous findings that he performed at chance level in a direct guessing task with static facial expression shown to his blind field (4). Although he was above chance for moving facial expressions, indirect testing methods provided evidence that still images are recognized (5).

The electrophysiological results obtained in GY and DB clearly indicate that facial expressions as well as emotional pictures influence the way emotional voices are processed as indicated by the decreased amplitude of the N1 component. The bias effect occurs early in the course of auditory perception (around 110 ms after the onset of the auditory stimulus). Our data indicate that intact striate cortex and conscious vision of the stimuli modulate this effect. Our results suggest that in the absence of V1, the crossmodal bias effect is restricted to the pairings of a voice with a face and does not obtain for the pairings of a voice with an emotional picture, although these pictures can reliably be discriminated in the blind field.

The N1 (or N100) auditory component is described as a late cortical exogenous component with a central topography, and is composed of multiple subcomponents (29). Its amplitude is mod-

ulated by auditory selective attention with an enlarged N1 elicited by attended stimuli (30). The present EEG results obtained for stimuli in the normal visual fields of two hemianopic patients are similar to previous EEG results obtained with normal observers showing that a facial expression modulates concurrent voice processing as early as 110 ms after voice onset (12, 13). The fact that the crossmodal bias effect consists in a decrease in amplitude of the auditory N1 component for incongruent face-voice pairs suggests that early voice processing is reduced in the context of an incongruent visual stimulus. Our results are consistent with reported sensitivity of the auditory N1 component to the audiovisual pairing (31). Visual bias of auditory processing has also been observed with functional MRI and consisted of an increase of the blood-oxygen level-dependent response in the primary auditory cortex when heard syllables were accompanied by lip reading (32). Moreover, a reduction in amplitude of voice processing (as indexed by the N1 component) as a function of the concurrent visual context is mainly observed for electrode positions located in the right hemisphere irrespective of the side of lesion of the patient. This observation is consistent with neuropsychological and brain-imaging data indicating the preferential involvement of the right hemisphere during the processing of emotional prosody (33).

We can discount three possible explanations for the observed difference in crossmodal effect between face pairs and scene pairs. First, it is unlikely that the observed effects are related to attentional factors or to any kind of response bias. Throughout the EEG experiment, subjects' attention was kept on a task-irrelevant property of the voice. Moreover, the fact that a facial expression influences concurrent processing of the voice expression as early as 110 ms after stimulus suggests that this pairing takes place at an early perceptual stage, in the sense that it does not depend on a postperceptual decision that might be under endogenous attentional control (10). A second potential explanation that we can discard is that the N1 modulation does not reflect crossmodal binding but instead reflects a heightened state of affect in the system caused by the simultaneous presence of the face or scene and the voice. This explanation would predict, contrary to our findings, a similar effect for the face-voice and scene-voice pairings. Third, the possibility that the difference in the crossmodal bias for faces and pictures in the blind field results from a difference in visual complexity is unlikely because the behavioral data indicate that faces and pictures are discriminated equally well by DB and pictures are actually better discriminated than faces by GY.

What are the possible anatomical circuits that could mediate audiovisual interactions and could explain why, in the absence of striate cortex, some pairs are preserved while others are lost? If faces and pictures can be equally well discriminated in the blind field, the differences in the results obtained when each is paired with the same voice may reside in the critical contribution of striate cortex and/or other cortical processes to one type of pairing but not to the other. Striate cortex loss may damage the mechanism that is the basis of scene-voice pairings but leave intact the neural circuitry for face-voice pairings. In this context, two alternatives can be envisaged at present. One relates to visual awareness *per se*. A striate cortex lesion is associated with loss of conscious vision, which might be assumed to be necessary for intersensory binding. Such a view is inconsistent with the finding that crossmodal effects can be covert (15). Beyond that, it fails to explain why there is a difference in binding between voices with scenes, on the one hand, and faces, on the other hand.

A second explanation can be proposed that focuses instead on limbic-cortical routes and the role of striate cortex in cortico-cortical feedback loops from more anterior areas. Such an explanation would run as follows. With striate cortex damage, other structures receiving a direct retinal input, such as the superior colliculus and the pulvinar, can compensate to some extent and support some visual discriminations, but they cannot compensate for feedback from anterior cortical areas to early visual areas (17,

34, 35). Residual vision based on such extrastriate routes probably lies at the basis of blindsight for faces and pictures alike, as observed in the present behavioral data. But the extent to which these routes can compensate in a multisensory setting appears to be limited. Colliculus- and pulvinar-based vision does not compensate for cortico-cortical connections and *a fortiori* does not compensate for cortico-cortical feedback projections in which the striate cortex is normally involved. Both types of connections might be needed for the involvement of vision in semantic pairings like the picture-voice ones used here. Although facial expressions and affective pictures can be successfully processed by observers with striate cortex damage, the former but not the latter can interact with information from another sensory system because the latter, unlike the former, requires intact cortico-cortical processing loops.

At present, support for this explanation is provided by the finding that noncortical vision of facial expressions is not associated with fusiform activity (6), by the finding of increased amygdala activation for face-voice pairs (11) and of increased activity in heteromodal cortical areas such as the middle temporal gyrus (G.P., B.d.G., A. Bol, and M. Crommelinck, unpublished data). The point is also illustrated in a functional MRI study with GY where we observed increased activation to fearful face-voice pairs in the amygdala and in the fusiform area for visual presentations in the intact field (6). The important finding, as far as the present explanation is concerned, is that amygdala activation was equally observed when faces were presented in the blind field but without the associated increase in activation in fusiform cortex. When affective pictures are recognized unconsciously, as our behavioral data indicate that can be, visual recognition may also be implemented via a route to limbic structures, but not achieve activation in a cortical area or areas necessary for perception of pictures, homologous to the involvement of fusiform cortex in perception of faces. This absence of cortical activation normally associated with subcortical activity may be a crucial element for understanding that noncortical routes can sustain some audiovisual pairings but not others. One may conjecture that in the neurologically intact observer anterior activation is

associated with feed-forward cortico-cortical activation from fusiform to heteromodal areas (35) (as well as feedback activation to striate cortex and conscious perception as envisaged in the first explanation above). Yet our results indicate that the probable nonstriate neural circuitry processing these face stimuli is sufficient for binding face-voice pairs. This finding suggests that the cortical activity that accompanies perception of a face expression in the intact visual field is not a crucial requirement for successful voice-face pairing. However, what holds for face-voice pairs may not hold for picture-voice pairs even if on their own faces and pictures are processed by the same neural circuitry. Without the kind of activation in extrastriate areas associated with striate vision, binding between pictures and voices may not be possible. The need for cortical activation would correspond to the fact that the pairing between a picture and a voice requires mediation by semantic systems of the brain and involves higher and more anterior areas. Given the relevance of facial expressions for human communication and, more importantly, the biological link between voice prosody and face expressions, the existence of specialized systems underlying a natural audiovisual pairing like a face-voice combination as compared with a pairing such as picture-voice that requires semantic mediation, has evolutionary value.

Finally, the special status of the face-voice pairings might be boosted by an additional mechanism. Normal subjects spontaneously imitate facial expressions even if these are followed by a rapid backward mask and are not consciously perceived (36). An interesting possibility is that binding of a face and a voice is mediated by such spontaneous imitation of the facial expression, a mechanism that is absent in scene-voice pairs. It would be of interest to record muscular changes in the facial pattern of subjects in these studies, for the blind as well as intact visual fields, which could provide additional evidence not only of the existence of processing of emotional material but a possible mediating role in crossmodal binding.

We thank GY and DB for their cooperation and patience during long hours of testing.

1. Esteves, F. & Ohman, A. (1993) *Scand. J. Psychol.* **34**, 1–18.
2. Morris, J. S., Ohman, A. & Dolan, R. J. (1998) *Nature (London)* **393**, 467–470.
3. Morris, J. S., Ohman, A. & Dolan, R. J. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1680–1685.
4. de Gelder, B., Vroomen, J., Pourtois, G. & Weiskrantz, L. (1999) *NeuroReport* **10**, 3759–3763.
5. de Gelder, B., Pourtois, G., van Raamsdonk, M., Vroomen, J. & Weiskrantz, L. (2001) *NeuroReport* **12**, 385–391.
6. Morris, J. S., de Gelder, B., Weiskrantz, L. & Dolan, R. J. (2001) *Brain* **124**, 1241–1252.
7. Murrain, E. A. & Mishkin, M. (1985) *Science* **228**, 604–606.
8. Málková, L. & Murray, E. A. (1996) *Psychobiology* **24**, 255–264.
9. de Gelder, B., Vroomen, J. & Teunisse, J.-P. (1995) *Bull. Psychonom. Soc.* **30**.
10. de Gelder, B. & Vroomen, J. (2000) *Cognit. Emot.* **14**, 289–311.
11. Dolan, R. J., Morris, J. S. & de Gelder, B. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 10006–10010.
12. de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M. & Vroomen, J. (1999) *Neurosci. Lett.* **260**, 133–136.
13. Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B. & Crommelinck, M. (2000) *NeuroReport* **11**, 1329–1333.
14. Vroomen, J., Driver, J. & de Gelder, B. (2001) *Cognit. Affect. Neurosci.* **1**, 382–387.
15. de Gelder, B., Pourtois, G., Vroomen, J. & Bachoud-Levi, A. C. (2000) *Brain Cognit.* **44**, 425–444.
16. Weiskrantz, L. (1997) *Consciousness Lost and Found* (Oxford Univ. Press, Oxford).
17. Lamme, V. A. (2001) *Acta Psychol.* **107**, 209–228.
18. Marzi, C. A., Tassinari, G., Aglioti, S. & Lutzemberger, L. (1986) *Neuropsychologia* **24**, 749–758.
19. LeDoux, J. (1996) *The Emotional Brain: The Myerious Underpinnings of Emotional Life* (Simon and Schuster, New York).
20. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. (1995) *The International Affective Picture System (IAPS): Photographic Slides* (Center for Research in Psychophysiology, Gainesville, FL).
21. Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J. & Dolan, R. J. (1996) *Nature (London)* **383**, 812–815.
22. Baseler, H. A., Morland, A. B. & Wandell, B. A. (1999) *J. Neurosci.* **19**, 2619–2627.
23. Barbur, J. L., Watson, J. D., Frackowiak, R. S. & Zeki, S. (1993) *Brain* **116**, 1293–1302.
24. Barbur, J. L., Ruddock, K. H. & Waterfield, V. A. (1980) *Brain* **103**, 905–928.
25. Weiskrantz, L., Warrington, E. K., Sanders, M. D. & Marshall J. (1974) *Brain* **97**, 709–728.
26. Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications* (Oxford Univ. Press, Oxford).
27. Ekman, P. & Friesen, W. (1976) *Pictures of Facial Affect* (Consulting Psychologists Press, Palo Alto, CA).
28. Weiskrantz, L., Cowey, A. & Hodiott-Hill, I. (2002) *Nat. Neurosci.* **5**, 101–102.
29. Nataanen, R. (1992) *Attention and Brain Function* (Erlbaum, Hillsdale, NJ).
30. Hillyard, S. A., Mangun, G. R., Woldorff, M. G. & Luck S. J. (1995) in *The Cognitive Neurosciences*, ed. Gazzaniga, M. S. (MIT Press, Cambridge, MA), pp. 665–681.
31. Giard, M. H. & Peronnet, F. (1999) *J. Cogn. Neurosci.* **11**, 473–490.
32. Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D. & David, A. S. (1997) *Science* **276**, 593–596.
33. Ross, E. D. (2000) in *Principles of Behavioral and Cognitive Neurology*, Mesulam, M. M. (Oxford Univ. Press, Oxford), 2nd Ed., pp. 316–331.
34. Pascual-Leone, A. & Walsh, V. (2001) *Science* **292**, 510–512.
35. Bullier, J. (2001) *Trends Cogn. Sci.* **5**, 369–370.
36. Dimberg, U., Thunberg, M. & Elmehed, K. (2000) *Psychol. Sci.* **11**, 86–89.