Seeing cries and hearing smiles

de Gelder, B.; Vroomen, J.; Pourtois, G.R.C.

Published in:
Cognitive contributions to the perception of spatial and temporal events.

Publication date:
1999

Cognitive Contributions to the Perception of Spatial and Temporal Events
G. Aschersleben, T. Bachmann, and J. Müsseler (Editors)

Chapter 17

Seeing Cries and Hearing Smiles:
Crossmodal Perception of Emotional Expressions[1]

Beatrice de Gelder[1,2], Jean Vroomen[1], & Gilles Pourtois[1,2]

[1]*Tilburg University, The Netherlands*
[2]*Neurophysiology Laboratory, Medical School, Louvain University, Belgium*

The perception of facial and vocal emotions is an essential part of the communicative competence upon which complex social interactions are based. The investigation of simultaneous emotional processing in more than one modality represents a new field of research for cognitive psychologists and neuroscientists alike who can start from these bimodal situations to study the behavioral and anatomical basis of this particular multimodal integration. The purpose of the present chapter is to present a review of the experimental work from a behavioural and electrophysiological perspective. We argue that integration of the emotion presented concurrently in the face and the voice is mandatory, automatic and not under attentional control. At the electrophysiological level, it produces an early negative component with a centro-frontal topography, entirely compatible with the properties of the well-studied mismatch negativity component. Audiovisual emotion perception certainly requires further research to bring some new lights on the questions raised by the results available these days.

1 Introduction

The two themes of this chapter, inter-sensory relations and perception of affective stimuli belong to domains of research that seem far and wide apart. Yet in each of these research domains action is a central concept. A persistent theme in the study of inter-sensory relations has been that of the crucial role of action in attuning different sensory systems. For example, in the seminal paper by Held and Hein (1958) examining the effect of a prismatic displacement of the visual field, adaptation is either attenuated or suspended altogether if the subject is not allowed self-initiated action. Action and movement are concepts that have played an important role since the first systematic attempts to describe emotions and perception of affective states (see Frijda, 1989, for an historical overview). The notion that perception is for action evokes the well-known view of William James (1884) on the relation between the perception of an emotional situation, the behaviour and the subjective awareness

[1] Thanks to P. Bertelson and to R. Held for discussions on crossmodal perception.

and experience of the emotional stimulus. Contemporary studies in affective neuroscience converge on the existence of two basic systems underlying different manifestations of emotion, an approach-related and a withdrawal-related one (see Davidson, 1998, for an overview). The present chapter reviews research that is at the crossroads of these two research domains and reports evidence from studies that have focused on the integration of the auditory and the visual modality in the processing of emotional messages. Facial and vocal emotional behaviors are an essential part of the communicative competence upon which complex social interactions in animal and man is based. How does the organism process multiple cues provided by different sensory modalities when these are present at the same time, like a facial expression and a tone of voice? Is there evidence that the voice and the face inputs are integrated and lead to a single percept or, alternatively, are the visual and the auditory percept combined in the course of a post-perceptual decision, after each has been fully processed separately?

2 Perceiving Emotions by Ear and by Eye

The situation where a voice and a face expression both signal an emotion is very familiar from everyday life. It is thus a bit surprising that the phenomenon has not until now caught the attention of laboratory researchers. We submit that one explanation for this state of affairs is the tacit assumption that information from the face expression and from the tone of voice are taken as interchangeable. In pondering whether somebody is angry, it usually suffices to either see or hear him. For all practical purposes, either seeing or hearing will do and the redundance does not seem to serve any particular purpose. This situation gives raise to the assumption that input from the eye and from the ear are both dealt with by the same processor and computed in an amodal representation system. Very few studies are available that have actually investigated the validity of that assumption. Walker and Grolnick (1983) studied the inter-modal perception of emotions in infants by presenting faces combined with voices. 5-7 month old infants looked longer at the face that carried the same expression than at the one carrying a different expression. Tartter and Braun (1994) studied the perception of the emotional meaning of syllables as a function of the facial expression the speaker had adopted in pronouncing them. They observed that a listener could gather the emotional expression of the face from listening to the syllables only. The finding that listeners can tell a speaker's emotional expression suggests that they might retrieve a production link between two separate inputs, the intonation of the voice and the expression on the face. Such an approach was defended in the earlier motor theory of speech perception (Liberman and Mattingley, 1985), and it is at the heart of some ecological approaches to understanding the processing of speech sounds.

A recent study by Massaro and Egan (1996) comes closest to the work we will describe in more detail below. These authors used a synthetic face combined with a spoken word and report evidence for the perceptual combination of both. In our own studies we have adopted an experimental situation that is familiar from the work by McGurk and McDonald (1976) showing that concurrently presented but incompatible

information provided by the lips and by the voice leads to an illusory percept. The paradigm that has been extensively used to examine the combination of auditory with visual information in speech perception is a variant of the categorical perception paradigm. In a number of experiments Massaro and collaborators have shown that adding lipread information has an impact on the location of the auditory identification curve. In a typical experiment, a visual stimulus (/ba/ or /da/) is combined each time with one of the stimuli of an auditory stimulus continuum obtained by stepwise synthesis between a /ba/ and /da/. The combination of a visual with an auditory syllable changes the way the auditory syllable is perceived. One observes a displacement of the identification curve to the right or the left depending on whether a visual /da/ or /ba/ is added. Cross-modal effects have also been observed for the combination of a still face and a voice (Campbell, 1996). The cross-modal bias effect is very robust and is obtained in a condition of integration where subjects are asked to repeat what the speaker says, as well as in a condition of selective attention where subjects are instructed to attend to the auditory information only.

Generating audiovisual conflict provides us with a paradigm that allows to pull apart the two processing streams and to create conditions for observing each separately as well as their interaction. In creating an experimental situation of the kind we will present below, subjects are instructed to combine what they hear with what they see. It is important to rule out that the obtained results represent an artefact created by the instructions and that the data would reflect the decision strategy the subject adopts in the presence of two different inputs of which he is separately aware rather than a mandatory integration of both in the course of perception. Our experiments were designed to rule out such an explanation.

3 Behavioral Evidence for Cross-Modal Effects between Voice and Face

In a series of recent experiments we tackled one aspect of bimodal processing concerned with the combination of the auditory and visual source and its effect on the latencies and the judgement of the displayed emotion. Our experiments used the paradigm of cross-modal bias. We first set out to examine the effect of a combination of a voice and face expression. The faces were taken from a continuum extending between two posed tokens expressing sadness or happiness. The two tones of the voice were also sad or happy. On each bimodal trial, a still photograph of a face was presented on the screen while a voice was heard pronouncing a sentence in one of two affective tones. Subjects were instructed to judge the emotion of the person (Experiment 1, de Gelder & Vroomen, 1995, 1999). We observed a huge difference in how the face was rated in the presence of either one of the two voices compared with the situation when only the face had to be recognized. Of particular interest are the bimodal trials. When the expression in the voice and that on the face are congruent, subjects are faster to judge the expression than when they only receive one input. In the case an ambiguous face is presented the rating of the face is strongly affected by the concurrent voice. The result appears to provide evidence for the combination of voice and face information in the course of processing the emotional content.

The question that is immediately prompted by such a result is whether such a combination will still take place when the subjects' attention is directed to one of the two sources. In other words, will this cross-modal effect resist an attention manipulation in which subjects are explicitly told to ignore one of the sources? It may indeed be argued that the task of judging the emotion generated by combining a still face and a spoken sentence is an artificial one and that the observed effects are due to the fact that the instructions are compelling rather than presenting any evidence for how the processing system tackles this situation. The instruction might have functioned as explicit cue to put together a voice and a face. The suspicion that the effect is due to instructions and depends on explicit attention to the two channels is reinforced by the fact that still faces were used. Since the same effect has been obtained with dynamic faces (de Gelder, Vroomen, & Weiskrantz, 1999).

The same experiment was thus repeated but with different instructions (de Gelder & Vroomen, 1995, Experiment 2). We now instructed the subjects to strictly judge the face only and to ignore any auditory information. The results showed that modifying the instructions did not change the basic pattern of results. We again observed a perceptual shift and noted that here also the voice has a significant effect on how the face was judged. Also, subjects are faster when they respond in the presence of two congruent inputs (happy face with happy voice) than when they are presented a face only.

Since the recognition of a facial expression is under the impact of a concurrently presented voice even if the perceiver is instructed not to pay any attention to the voice, we may conclude that the combination of audition and vision is automatic. Such automatic effects or mandatory phenomena are contrasted with post-perceptual effects like the ones that result from a subjective decision.

Does the reverse effect also obtain: Can processing of voice be affected by concurrently presented visual information? We examined this question by designing a situation similar to the previous one where the task was to judge the expression in the voice ignoring the information concurrently conveyed by the face. The answer to this question is clearly positive (de Gelder & Vroomen, 1999, Experiment 3).

The observed phenomenon of cross-modal bias in the perception of emotion by ear and by eye is particularly striking because it is obtained in a situation that does not mimic the natural situation. In fact, our experimental situation only resembles very superficially the natural, ecological situation of concurrent inputs. Normally in real life the face and the voice express the same emotion but in our paradigm a mismatch was created artificially. The interesting finding is that the system is strongly biased towards putting together information from voice and face. Subjects are sometimes aware of an inconsistency between the voice and the face expression, but this phenomenal impression of inconsistency between the two sources seems to belong to a different, possibly higher and conscious level of processing that does not interfere with the compelling bias for the processing system.

The direction indicated by our studies is that information from seeing and hearing the voice combine early. But it is important to note that this result is still compatible with the notion that processing of emotion in the face and in the voice is carried out in different, modality specific representation systems. Integration takes place after the respective sensory sources have been fully processed (late integration models).

Such an approach has similarities with the standard late integration view of the Stroop effect (McLeod, 1991). A different approach to audiovisual theories is the postulated recoding of the input representation. Either one source is recoded into the representational system of the other or both sensory representations are recoded into a supra-modal or abstract representation system. A third possibility is that of parallel extraction of information in the two modalities. A version of this view has been defended over the last decades by Massaro, whose model is intended to be applicable to any situation of concurrent inputs and more generally, any situation of multiple information sources (see most recently Massaro & Egan, 1996).

The development of a proper theory of intersensory emotion perception requires further research than is presently available. Two major issues are that of the time course of the audiovisual combination at the basis of the cross-modal bias and that of the domains of information that do interact. As to the former issue, our reaction time data as well as the robustness of the bias effect with the attention manipulation are compatible with the notion of an early integration. But a full answer to the question on the time course of integration may require evidence from other than strictly behavioral methods as argued by Stein and Meredith (1993). This issue brings up the next question.

4 The Time Course of Audiovisual Emotion Perception: An Electrophysiological Study

When does the mind/brain put together what it hears and what it sees? An answer to this question has been pursued by cognitive psychologists applying detailed chronometric methods and have made the case that we are dealing with truly perceptual phenomena where combination of the input streams is mandatory, not reflecting a perceptual bias or a postperceptual decision process under subjective control. The observation of shorter latencies with congruent voice-face combinations over presentations of the face only indicates that it is somehow more efficient for the system to receive bimodal input. What functional and neuroanatomical model underlies this apparent gain? One possibility is that inputs are combined early on and that the combination allows for a faster percept. But this does not need to imply that the inputs are actually combined, not a fortiori that there are common processing resources or shared representations for faces and voices. Shorter latencies with congruent bimodal representations are compatible with a race model, that is, the input that is processed fastest determines the outcome. The matter is hard to settle with behavioral data only.

The method of recording Event-Related Potentials (ERPs) allows us to address issues of the time course of cognitive processes at stake in understanding language. It has been used to study recognition of face processing and of voice expressions. But the combination of voice and face expressions has so far not been the focus of any study. Yet the tools are there to address this issue. There exists an ERP component which is known to be only sensitive to one modality but might be very useful in the study of combined inputs also, the mismatch negativity (MMN). It is known to reflect processing of auditory stimuli (see Näätänen, 1992, for an overview). The

MMN is elicited by a deviant stimulus in a repetitive train of standard auditory stimuli. It is an autonomous brainwave not controlled by attention and its amplitude is larger for larger differences between the standard and deviant stimuli as well as for subjects who show a greater sensitivity to that change at the behavioral level. Recently we exploited the potential of the MMN for tracing the early combination of the affective tone of the voice with information provided by the expression of the face (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; Pourtois, de Gelder, Vroomen, Rossion, Guerit, & Crommelinck, 1999). Subjects receive concurrent voice and face stimulation, but the face sometimes carries an incongruent emotional expression (as realized in the McGurk effect for speech). If the system is tuned to combine this dual input, as was suggested by our behavioral experiments, and if this combination consists in an early influence of the face input on the processing of the voice, this will be reflected in an MMN, or other auditory ERP components. If not, only the visual ERP components will be affected.

Our results indicate that when after a number of presentations of a voice-face pair both with the same expression, a pair is presented where the expression of the face is different, an early (100 - 200 ms) ERP component is elicited which strongly resembles the MMN typically associated with detection of a change in the auditory input. The distribution (Fz maximum) and latency (178 ms) are entirely compatible with those of the MMN (Näätänen, 1992).

The major theoretical importance of the ERP study consists in the direct evidence that face and voice input are combined early, at the latest at 178 ms after voice onset. This corroborates the conclusion based on behavioral research which supported an early combination of face and voice expressions. The behavioral data show that the identification of the voice expression is indeed hampered by a simultaneously presented incongruous face. But the present electrophysiological evidence that both inputs are combined at an early stage seems to rule out one of the alternatives left open by the behavioral results, that of a race model between separately processed faces and voices.

5 Evidence from Neurologically Impaired Patients

Important evidence for modality-specific representation systems that has been found in the domain of speech comes from dissociations in patients suffering from neuropsychological disorders. Selective disruption of speechreading with preserved ability to process auditory speech is a prime example (de Gelder, Vroomen, & Bachoud-Levi, 1998). Similar evidence is beginning to emerge for emotion perception. Recent data from brain damaged subject suggests that a deficit in face perception doesn't impair recognition of facial expression. More intriguingly, in some cases explicit recognition of facial expression is lost but the patient continues to process facial expressions without however being aware of doing so. Such covert recognition of facial expressions was observed when faces were presented on their own but also when the impact of a face expression on the voice was studied. De Gelder, Vroomen and Bachoud-Levi (1997) studied a prosopagnosic patient AD who was unable to recognize any emotion from still faces. We presented her with a version of the

audiovisual experiments described above and noted that the expressions she could not recognize in isolation nevertheless had an impact on her recognition of the voice expression. The literature on preserved covert recognition of faces offers a cue as to the routes that might be involved in this case of implicit expression recognition. A similar observation was made for a very different kind of patient, GY, suffering from blindsight. GY could reliably tell apart two facial expressions presented to his blind field (de Gelder, Vroomen, & Weiskrantz, 1999).

6 Cognitive Factors in Intersensory Bias

A central theme of our presentation of the research on audiovisual perception of emotions is that in witnessing the impact of a face on a voice and vice versa, we are dealing with a truly perceptual as opposed to a cognitive phenomenon. At the most general level this amounts to the claim that the integration of audition and vision is driven by perceptual principles and does not depend on cognitive factors. Dependency on cognitive factors has variously been referred to as a matter of top-down influences or of post-perceptual effects or of subjective beliefs, or of cognitive penetrability. The strongest statement of such a perceptual as opposed to a cognitive position is to say that audiovisual integration is not cognitively penetrable (see Pylyshyn, in press, for a historical overview and recent arguments). If audiovisual perception of emotion is truly perceptual, our findings belong to a family of phenomena whose more established members include ventriloquism and the McGurk effect. In order to better understand to what extent this is indeed the case, we must confront our phenomenon with alternative, non-perceptual explanations, that is, explanations relating the phenomenon to cognitive factors. Research on intersensory bias has mainly discussed the possible impact of three cognitive factors: familiarity of the association, impact of instructions and the role of attention (Bertelson, 1999; Welch, 1999). We comment first on each of these. In the next section we discuss a notion of cognitive factors that is of a different order and relates to the content, the semantic aspects or the stimulus domains to which the inputs belong.

As noted by Bertelson and Welch it is particularly important for multisensory integration experiments to be clear about the role of higher order cognitive factors. In many situations of bimodal input subjects are at least to some extent aware of a discrepancy between the two sources and could develop a strategy of selectively attending to either the one or the other. Such a strategy might even be a consequence of subjective dominance of the perceivers for either audition or vision. A recent study by Giard and Peronnet (in press) illustrated this point. The researchers separated their subjects into two groups according to their dominant modality to perform the unimodal tasks. Electrophysiological results clearly show that the integration effects affected predominantly the sensory areas of the non-dominant modality.

The role of attention on intersensory bias has been a topic of some controversy. As discussed by Welch (1999) there is evidence indicating that the degree to which each sensory modality biases the other is related to the amount of attention a subject allows to that specific sensory modality. As we mentioned above, one of our experi-

ments addressed the issue of attentional distribution. In the experiments where subjects had to judge the voice in the presence of a face they were given a concurrent task which required processing of a visual stimulus (a digit) other than the face. The goal was to see whether attention to this irrelevant visual stimulus would reduce the impact of the visual information from the facial expression on the voice. The results showed that such is not the case and that interference with visual processing of the face from a concurrent task requiring visual attention does not reduce crossmodal bias. One might object though that a concurrent visual task, even if unrelated to the perceptual process under scrutiny, nevertheless indirectly enhances processing of the face expression because it allocates extra attention to the visual modality. Further research is needed to sort out whether the crossmodal effect of the unattended face is affected by an attention demanding task that is not about a visual stimulus.

A second cognitive factor that has long been suspected of playing a role in intersensory processes concerns instructional variables. Researchers generally agree that instructions are a possible contaminant of subjects' responses. The role of instructional variables has been addressed in our experiments by contrasting two different sets of instructions. De Gelder and Vroomen (1999, Experiment 1 and 2) compared the data from an experiment where the subject was told to judge the emotional state of the person with the data from the same experiment using the same design but requiring a judgement of the face expression. Although the size of the cross-modal bias effect is smaller in the latter Experiment, the ignored information from the voice continues to have a sizable significant effect on judging the emotion in the voice. The results from the electro-physiological study are particularly instructive as they provide evidence for a cross-modal bias effect that is too early in perception and too basic (it occurs before recognition of the stimulus) to be under instructional control.

Familiarity with the fact that a given visual stimulus is frequently if not always paired with an auditory input is a third factor that might influence the subjects willingness to opt for the unity assumption. In that case the strength of a pairing would be a function of the strength of the association between the two sources. If acquaintance with the stimulus pair or familiarity were to be an important factor in audiovisual pairings of sources of affective content, it would mean that two stimuli are processed as belonging to a single event against the background of such subjective knowledge and background expectations. The role of familiarity might be more pervasive and more difficult to control for in the case of the present phenomenon. Just like in the case of speech, the association of a smiling face with a happy voice is obviously a familiar one for the perceiver. In that sense, one might argue that our experiments only bring to light the obvious, which is that subjects know that these two sources are commonly related. However, our studies also have a contrasting condition where the subject is presented with a voice-face pair that do not occur together in natural ecological circumstances. If crossmodal bias were due only to familiarity of the association there should be no incentive or no compelling reason for the subject to assume a single event in such cases. Instead we observe that in the highly artificial situation where for example an angry voice is heard together with a happy face, the voice information is still processed and as a consequence the face

expression is judged to be less happy. The simplest version of such experiments is where we looked only for an effect in the naming latencies for the face expression and observed that a voice expression unrelated to the face expression task has an effect on the expression naming latencies (de Gelder *et al.*, 1999, Experiment 1)

7 Cognitive Factors and Semantic Content

So far we have discussed familiarity in the sense of subjective acquaintance with objectively co-occuring events in different modalities. A somewhat different meaning of familiarity is at stake when the case can be illustrated with speech gestures. Work in progress investigates the combination of a female face expression with a male tone of voice This will allow us to estimate the importance of semantic factors like gender consistency focus on the role of familiarity in the cross-modal association link.

When factors like familiarity, instructions and attention are considered to belong to cognitive variables the contrast envisaged is between sensory factors. When the role of cognitive factors was first discussed in the context of intersensory bias, the question was how to isolate the intersensory pairing mechanism from subjective beliefs, expectations and cognitions. Over the last decades we have witnessed the raise of a different notion of cognitive factors that relates to the semantic content of the processes. This is most clearly the case in models of information processing that consist of three layers. Between sensory processing on the one hand and higher cognitive processes on the other, attention has been drawn to an intermediate level, that of functional content processes, located in between sensory processes and higher order cognitions. Like sensory processes, the functional stages of information processing are automatic, mandatory and not open to introspection. Like higher order conscious cognitions they are sensitive to the content being processed and will selectively process only what belongs to their domain (*e.g.*, speech, faces, emotions). The combination of mandatory processing and context sensitivity is what characterizes this level of processing. Crossmodal processing of emotions like that of speech belongs to this level. The kind of processing of emotional meaning which is at stake in the studies just mentioned fits the notion of a modular process. The latter can be viewed as perception based as well as mandatory and has sometimes been assimilated to a cognitive reflex (Fodor, 1983). But these may not be the main nor the only properties by virtue of which a process qualifies as modular. Another important aspect of modular processing concerns semantics or the representational content implicated at the level of modular processing of emotional messages. An example from language processing illustrates the point. In listening to a spoken sentence containing the word 'bank', the language module parses the input and recognizes the word 'bank'. However, it does not actually select which of the different meanings of the word is at stake in the actual context in which the sentence is spoken. In other words, content is shallow and not integrated within the full belief and thought systems of the subject. One way to bring out this contrast is by opposing shallow vs. full processing of a stimulus (Fodor, 1983). Along these lines, one may view the representational content of emotional experience as only a matter of narrow, functional or modular content on the one hand and phenomenal content on the other, or

to paraphrase, between perceptual states of emotion and full blown elaborate and reflexive belief states, where the meaning of an emotional stimulus is elaborated against the full richness of the subjective experience. Perceiving emotions in the voice, the face or in both is a process that can bypass consciousness. This suggestion implies that the pairing mechanism at the basis of these effects is not so much or not exclusively based on sensory principles like spatio-temporal contiguity or similar Gestalt principles. Commonality of content is likely to be a relatively more important determinant to trigger the pairing mechanism. Future research should investigate how sensory mechanisms of pairing like for example the ones at work in ventriloquism interact and combine with content based mechanisms of the kind exemplified in bimodal speech and emotion processing.

8 Emotions and Awareness

If intersensory phenomena including the one we have discussed here, are truly perceptual, the three cognitive factors discussed above appear as so many sources of contamination of the perceptual basis of the effect. One very radical way of addressing that objection makes the situation completely non-transparent for the subject by designing the experiment such that the perceiver is not aware of the second concurrent input.

There is now increasing evidence that processing of emotional messages takes place outside the scope of awareness. When subjects rate a face for gender they nevertheless appear to fully process its emotional content (Morris *et al.*, 1996). Faces that are not perceived consciously nevertheless lead to activation of the amygdala (Whalen *et al.*, 1998). Patients that are unable to consciously report face expressions do nevertheless show evidence of having processed these expressions covertly. The studies on perceiving emotions by ear and by eye we summarized here are consistent with evidence that emotional messages are processed outside subjective awareness. The merging of the two input channels that convey emotional information in the case of a bimodal emotion event is thus achieved in an automatic fashion bypassing any conscious awareness including awareness of incongruence between the expression in the voice and that in the face.

It appears then that our common sense understanding about the richness of emotional experience contrasts with the evidence accumulating from the cognitive neurosciences that a significant part of emotional processes bypasses our subjective access to and our accountability for what we experience. Ledoux (1996) clearly illustrates how separate processing streams in the brain correspond to implicit and explicit emotion processes.

References

Ahern, G. L., & Schwartz, G. E. (1985). Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia, 23(6)*, 745-755.

Austin, J. L. (1970). Other minds. In J. O. Urmson & G. J. Warnock (Eds.), *Philosophical papers* (2nd ed., pp. 47-63), London: Oxford University Press.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Research, 342*, 91-102.

Bertelson, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, F. I. M. Craik and M. Robert (Eds.) *Advances in Psychological Science, Vol.1: Biological and Cognitive Aspects* (pp. 419-439). Hove, UK: Psychology Press.

Bertelson, P. (1999). Ventriloquism: Cross-modal patterning under spatial conflict. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*. Amsterdam: Elsevier (this volume).

Bertelson, P., Vroomen, J., de Gelder, B. & Driver, J. (1999). *Auditory-visual spatial interaction and the orientation of visual attention.* (submitted manuscript)

Bower, T. G. R. (1974). *Development in infancy*. San-Fransisco: W.H. Freeman.

Bruyer, R. (1981). Asymmetry of facial expression in brain damaged subjects. *Neuropsychologia, 19(4)*, 615-624.

Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. L. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology, 87(1)*, 141-162.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Campbell, R. (1996). Seeing speech in space and time. *Proceedings of the International Conference on Spoken Language Processing*, pp. 1493-1498.

Campbell, R., Dodd, B., & Burnham, D. (1998). *Hearing by Eye II: advances in the psychology of speechreading and auditory-visual speech*. Hove: Psychology Press.

Carmon, A., & Nachshon, I. (1973). Ear asymmetry in perception of emotional non-verbal stimuli. *Acta Psychologica, 37(6)*, 351-357.

Crowder, R., & Surprenant, A. (1995). On the linguistic module in auditory memory. In de Gelder, B. & Morais, J. (Eds.), *Speech and reading: A comparative approach* (pp. 94-64). Hove: Erlbaum.

Darwin, C. (1872). *The expressions of emotion in man and animals*. London: John Murray.

Davidson, R. J., & Tomarken, A. J. (1989). Laterality and emotion: An electrophysiological approach. In F. Boller and J. Grafman (Eds.), *Handbook of neuropsychology*, Vol 3 (pp. 419-441). Amsterdam: Elsevier.

de Gelder, B., Bachoud-Levi, A. & Vroomen, J. (1997). Emotion by ear and by eye: Implicit processing of emotion using a Cross-modal approach. Boston, March 23-25. *Fourth Annual Meeting of the Cognitive Neuroscience Society, 49*, 73.

de Gelder, B., Teunisse, J.-P., & Benson, P. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11*, 1-23.

de Gelder, B., & Vroomen, J. (1996). Categorical perception of emotional speech. *The Journal of the Acousical Society,* 100, 4, Pt. 2, 2818.

de Gelder, B., & Vroomen, J. (1994). Memory for consonants versus vowels in heard and lipread speech. *Journal of Memory and Language, 31*, 737-756.

de Gelder, B. & Vroomen, J. (1999). *Perceiving emotions by ear and by eye.* (submitted manuscipt).

de Gelder, B., & Vroomen, J. (1999). *Categorical perception of speech prosody.* (manuscript in preparation)

de Gelder, B., Vroomen, J. & Bertelson, P.(1998). Upright but not inverted faces modify the perception of emotion in the voice. *Current Psychology of Cognition, 17*, 1021-1031.

de Gelder, B., Vroomen, J., & Bachoud-Lévi, A. (1998). Impaired speechreading and audio-visual speech integration in prosopagnosia. In R. Campbell, B. Dodd, & D. Burnham

(Eds.), *Hearing by Eye 11, Advances in the Psychology of Speechreading and Auditory-visual Speech* (pp. 195-207). Hove: Psychology Press LTD.

de Gelder, B., Vroomen, J., & Bertelson, P. (1996). Aspects of Modality in audio-visual processes. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by Humans and Machines. NATO ASI Series F, Vol. 150* (pp. 197-192). Berlin: Springer-Verlag.

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensens, M. and Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by electric brain responses. *Letters in Neuroscience, 260,* 133-126.

Delis, D. C., Kiefner, M. G. & Fridlund, A. J. (1988). Visuospatial dysfunction following unilateral brain damage: Dissociations in hierarchical hemispatial analysis. *Journal of Clinical and Experimental Neuropsychology, 10(4)*, 421-431.

Dodd, B., & Campbell, R. (1987). *Hearing by Eye: the psychology of lipreading.* Hove, UK: Lawrence Erlbaum.

Dopson, W. G., Beckwith, B. E., Tucker, D. M., & Bullard-Bates, P. C. (1984). Asymmetry of facial expression in spontaneous emotion. *Cortex, 20(2)*, 243-251.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54(3)*, 414-420.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169-200.

Etcoff, N., & Magee, J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227-240.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge: MIT Press.

Frijda, N. (1989). *The Emotions.* Cambridge: CUP.

Fried, I., Mateer, C., Ojemann, G., Wohns, R., & Fedio, P. (1982). Organization of visuospatial functions in the human cortex. *Brain, 105*, 349-371.

George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., Herscovitch, P., & Post, R. M. (1996). Understanding emotional prosody activates right hemisphere regions. *Archive of Neurology, 53*, 665-670.

Giard, M. H., & Peronnet, F. (in press). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience.*

Gainotti, G. (1972). Emotional behavior and hemispheric side of lesion. *Cortex, 8*, 41-55.

Gainotti, G. (1989). Disorders of emotions and affect in patients with unilateral brain damage. In F. Boller and J. Grafman (Eds.), *Handbook of Neuropsychology, Vol. 3* (pp. 345-361). Amsterdam: Elsevier.

Gianotti, M., & Pereira, N. (1972). Psychological approaches to the cerebral palsied child. *Revista Brasiliera de deficiencia mental, 7(1)*, 5-9.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Humphreys, G. W. and Riddoch, J. M. (1987). *To see or not to see.* Hove: Lawrence Erlbaum.

James, W. (1884). What is an emotion? *Mind, 9*, 188-205.

Johnson, W. F., Emde, R., Scherer, K., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry, 43*, 280-283.

Landis, T., Assal, G., & Perret, E. (1979). Opposite cerebral hemispheric superiorities for visual associative processing of emotional facial expressions and objects. *Nature, 278*, 739-740.

Lane, R. D., Reiman, E. M. Geoffrey, L. A., Schwatrz, G. E. and Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness and disgust. *The American Journal of Psychiatry, 154*, 926-933.

Ledoux, J. (1996). *The emotional brain.* New York: Simon and Schuster.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development, 8*, 335-353.

Leonard, C. M., Rolls, E. T., Wilson, F. A., & Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research, 15(2)*, 159-176.

Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: MIT Press.

Liberman, A. M. and Mattingley, G. (1993). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to emotional content of speech. *Journal of the Acoustical Society of America, 34*, 922-927.

Locke, J. (1689). An essay concerning human understanding. Repr.J.W. Yolton (Ed.), 2 vol.,1967-68, London : Dent.

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

Mahoney, A. M., & Sainsbury, R. S. (1987). Hemispheric asymmetry in the perception of emotional sounds. *Brain and Cognition, 6(2)*, 216-233.

Massaro, D. W. (1987). *Speech perception by ear and eye.* Hillsdale, NJ: Lawrence Erlbaum.

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

McGurk, H., & MacDonals, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meltzoff, A. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory. *Annals of the New-York Academy of Sciences, 608*, 1-37.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*, 1097-1108.

Näätänen, R. (1992). *Attention and brain function.* Hillsdale, NJ: Lawrence Erlbaum.

Nahm, F. K., Tranel, D., Damasio, H., & Damasio, A. R. (1993). Cross-modal associations and the hyman amygdala. *Neurospychologia, 31(8)*, 727-744.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Pollack, I., Rubenstein, H., & Horowitz, A. (1960). Communication of verbal modes of expression. *Language and Speech, 3*, 121-130.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Guerit, J.M., and Crommelinck, M. (1999). *Multimodal perception of emotion: an event-related potentials study.* (submitted manuscript).

Pylyshyn, Z. (in press). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences.*

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the hunam auditory cortex. *Neuroscience Letters, 127*, 141-145.

Scherer, K. R. (1979). Non-linguistic vocal indicators of emotions and psychopathology. In C. E. Izard (Ed.), *Emotions in personality and psychopathology.* Plenum Press, New York.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J. *et al.* (1997). Impaired auditory recognition of fear and anger following bilateral amygda la lesions. *Nature, 385(6613)*, 254-275.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22(4)*, 904-915.

Stein, B. E. & Meredith, M. A. (1993). *The merging of the senses.* Cambridge, Mass: MIT Press.

Suberi, M., & McKeever, W. F. (1977) Differential right hemispheric memory storage of emotional and non-emotional faces. *Neuropsychologia, 15(6)*, 757-768.

Summerfield, Q. (1992). Towards understanding audiovisual speech recognition. In G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 117-137).

Tartter, V., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96*, 2101-2107.

Teunisse, J. P., & de Gelder, B. (in press). Impaired categorical expression of emotions in autistics. *Child Neuropsychology.*

van Lancker, D. (1997). Rags to riches: Our increasing appreciation of cognitive and communicative abilities of the human right hemisphere. *Brain and Language*, 57, 1-11.

van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35(5)*, 963-970.

Vroomen, J., Collier, R., & Mozziconacci, S. (1993). *Duration and intonation in emotional speech.* Proceedings of the Third European Conference on Speech Communication and Technology, Berlin, 577-580.

Vroomen, J., & de Gelder, B. (1996). Phoneme detection in resyllabified word. *The Journal of the Acoustical Society.* 100, 4, Pt. 2, 2818-2819.

Walker, A., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6*, 491-498.

Weiskrantz, L. (1997). *Consciousness Lost and Found.* Oxford: Oxford University Press.

Welch, R. B. (1999). Meaning, attention and the "Unity Assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events.* Amsterdam: Elsevier (this volume).

Werner, H. (1973). *Comparative psychology of mental development.* New-York: International Universities Press.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: some acoustic correlates. *Journal of the Acoustical Society of America, 52*, 1238-1250.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*,141-145.