Kaufmann, Jurgen Michael (2002) *Interactions between the processing of facial identity, emotional expression and facial speech?*
PhD thesis.

INTERACTIONS BETWEEN THE PROCESSING OF FACIAL IDENTITY, EMOTIONAL EXPRESSION AND FACIAL SPEECH?

Jürgen Michael Kaufmann

University of Glasgow

Department of Psychology

Scotland, U.K.

Submitted for the degree of Ph.D. to the Higher Degree Committee of the Faculty of Social Sciences, University of Glasgow

May, 2002

Acknowledgements

These studies were supported by the Department of Psychology, University of Glasgow, Scotland and the Deutsche Forschungsgemeinschaft (Schw 511/4_1).

The help of Anna Landgrebe in data collection and stimulus editing, and the help of Marco Steinhauser in programming and data collection is gratefully acknowledged. I also would like to thank Wolfgang Hesseler for adaptations of the QuickView^{TM} software. Special thanks to my supervisor Prof. S.R. Schweinberger for his support and valuable comments on a previous draft. Finally, I would like to thank Lucila and our families for all the support and encouragement.

Abstract

The experiments investigate the functional relationship between the processing of facial identity, emotional expression and facial speech. They were designed in order to further explore a widely accepted model of parallel, independent face perception components (Bruce and Young, 1986), which has been challenged recently (e.g. Walker et. al., 1995; Yakel et al., 2000; Schweinberger et al., 1998; Schweinberger et al., 1999). In addition to applying a selective attention paradigm (Garner, 1974; 1976), dependencies between face related processes are explored by morphing, a digital graphic editing technique which allows for the selective manipulation of facial dimensions, and by studying the influence of face familiarity on the processing of emotional expression and speechreading. The role of dynamic information for speechreading (lipreading) is acknowledged by investigating the influence of natural facial speech movements on the integration of identity specific talker information and facial speech cues.

As for the relationship between the processing of facial identity and emotional expression, overall the results are in line with the notion of independent parallel routes. Recent findings of an "asymmetric interaction" between the two dimensions in the selective attention paradigm, in the sense that facial identity can be processed independently from expressions but not vice versa (Schweinberger et al., 1998; Schweinberger et al., 1999) could not be unequivocally corroborated. Critical factors for the interpretation of results based on the selective attention paradigm when used with complex stimuli such as faces are outlined and tested empirically.

However, the experiments do give evidence that stored facial representations might be less abstract than previously thought and might preserve some information about typical expressions. The results indicate that classifications of unfamiliar faces are not

influenced by emotional expression, while familiar faces are recognized fastest for certain expressions.

In contrast to previous reports of influences of speaker variations on statically presented familiar and unfamiliar faces (Schweinberger et al., 1998), effects of irrelevant speaker variations on speechreading speed appeared to be largely restricted to dynamic facial speech and unfamiliar speakers. The results underline the crucial role of dynamic information for speechreading (see also Rosenblum et al., 1998). They also provide evidence that speechreading from moving faces might be faster when these are familiar, which might point into the direction of an early and rapid integration of identity and dynamic facial speech information.

The results were discussed in the context of a functional model of face perception (Bruce and Young, 1986), assuming strictly modular processing of identity, expression and facial speech, and a more recent distributed neural model of face perception (Haxby et al., 2000) which takes into account the possibility of interactions between the brain structures which are now widely assumed to play a major role for face perception.

Table of contents

List of illustrations

1 Introduction

1.1 Topic and approach

The human face is a uniquely rich social stimulus, revealing a variety of social information such as age, gender, identity and the emotional state of an individual. Almost without effort we distinguish unknown from familiar faces and often we are successful in retrieving semantic information and the name belonging to a face. At the same time we can make inferences about the emotional state of a person by interpreting the activation pattern of facial muscles, and we can decide whether we want to approach an individual or prefer to stay out of reach. Faces may also influence the attribution of personality traits. Finally, through dynamic variations in mouth shapes during speaking, faces communicate visual speech information that has been shown to contribute significantly to speech perception (e.g. Miller et al., 1955; Sumby et al., 1954).

The relationship between these components of human face perception is yet not completely clear. According to an influential model of face processing (Bruce & Young, 1986), specialized modules for face recognition, emotional expression and speechreading work in parallel and are independent from each other's output (see also Ellis, 1989; Parry et al., 1991). Most studies in the field are restricted only to one aspect of face perception and there is a clear cut in the literature between research dealing with either face recognition, the processing of expression or facial speech. It can therefore be said that the independence between the processing of emotional expression, facial identity and speechreading has often been claimed but rarely been clearly demonstrated. It might well be that the independence between these processes suggested by some nil results in some experimental studies (e.g. Etcoff, 1984) rather reflects insufficiently rigorous testing than strict functional independence. Most of the evidence for an independent processing of facial identity, expressions and speechreading comes from neuropsychological studies that show selective impairments and double-dissociations between particular functions (Bornstein, 1963; Campbell et al., 1986; Kurucz et al., 1979a; Kurucz et al., 1979b; Parry et al., 1991; Young et al., 1993). These findings suggest different anatomical substrates for various aspects of face perception. However, different substrates and relative sparing of functions after focal brain damage do not rule out the possibility that under normal conditions functional units interact with one another, and that such interactions might improve face processing in healthy individuals. In a cascade model, such interactions are possible on any processing stage, and damage to one modality can affect other functions in various ways. Indeed, most neuropsychological studies demonstrating impaired processing of either identity, expression or speechreading after brain damage, also show some impairment of "intact" functions and might therefore argue for a weak independence theory. Although correlations between impaired functions can be caused by widespread lesions affecting different modules, it cannot be completely ruled out that such a pattern could also be due to impaired interactions between neural units.

Why should there be a cross talk between the processing of facial identity, emotional expression and facial speech? Depending on language, regional dialects, social status and language problems such as stuttering or lisps, but also as a result of normal individual variation (Montgomery et al., 1983) there are considerable inter-individual differences in the pronunciation of speech utterances. All of these are also visible in the dynamically changing mouth, jaw and lip patterns that occur while speaking. A system that is able to take such idiosyncratic differences into account might be more efficient in processing facial speech, so that it should have an ecological advantage. It has to be stated however, that speechreading certainly cannot be completely dependent on facial identity, since it is perfectly possible to speechread from unfamiliar faces. A flexible speechreading system, which is largely independent of supposed identity modules, but which is able to use their output when such idiosyncratic information is present, is in line with relatively circumscribed impairments after focal brain damage. A similar argument is possible for the processing of emotional expressions, although it might perhaps be less convincing. People differ in the way they display emotions, but expressions change face shapes according to regular, generalizable principles across cultures (Ekman, 1982). Also, the restricted number of basic emotions contrasts with the large number of possible speech sounds and dialectal variations, possibly making it less necessary to adjust as quickly to idiosyncratic characteristics as it might be for speechreading.

In this study I aim at exploring possible *interactions* between the processing of facial identity, emotional expressions and facial speech in order to test the model of parallel independent processing. This work was mainly inspired by some recently published data, which challenge the view of a completely independent processing (e.g. Schweinberger et al., 1998; Schweinberger et al., 1999; Walker et al., 1995; Yakel et al., 2000). It is the aim of this study to add to the clarification of the relationship between the processing of facial identity, emotional expressions and facial speech in healthy individuals on a functional level, and further explore recent findings that contradict the notion of strictly independent processes in face perception.

1.1.1 Organisation

Because the experiments described here were mainly designed with respect to the model by Bruce and Young (1986) I will briefly outline this model before presenting empirical evidence. This will be followed by an overview of some of the relevant literature on the processing of facial identity, emotional expression and facial speech. The aim is not to give a complete overview of the research on each particular function, as this would go far beyond the scope of the topic, but I will try to delineate the relevant features that each supposed module depends on, and where possible, summarize some of the available evidence concerning associated anatomical substrates. Following that I will present evidence for and against the independence model. In section 1.3, I will explain the underlying rational for the experiments, especially with respect to the selective attention paradigm (Garner, 1974, 1976) and provide some information on the morphing technique, which was used for the selective manipulation of identity and expression in Experiments 8-12. Finally, chapters 2 to 7 are dedicated to empirical evidence.

1.2 Literature review

I will try to give a basic overview over findings on face recognition, the processing of emotional expressions and speechreading. For each function the relevant facial features, evidence for functional processing stages and associated neural structures will be outlined. After describing a recent attempt to integrate a variety of findings into a "distributed human neural system for face perception" (Haxby et al., 2000), I will give an overview over the evidence arguing for and against an independent parallel processing of identity, expression and facial speech.

1.2.1 The parallel model of face perception

Bruce and Young (1986) suggested a theoretical framework for the processing of faces. Their model, which is related to other functional models of face processing (see also Ellis, 1986; Hay et al., 1982) has proven highly influential. It has been refined since (e.g. Burton et al., 1990) and is able to explain a range of empirical findings. It assumes distinct functional modules as bases for independent parallel processes underlying face perception (see also Figure 1). Overall, the model makes more detailed assumptions about face recognition than about the processing of expressions or speechreading. Importantly, specialized modules are assumed to underlie each particular process. These modules are supposed to work in parallel and independently from each other's output (see also Ellis, 1989; Parry et al., 1991). All subsequent processes have the first stage in common: the *structural encoding level* provides descriptions, which form the basis for parallel and independent routes that deal with the processing of emotional expressions, the recognition of familiar faces, facial speech and directed visual processes. Each route is characterized by a hierarchical and sequential processing, but both top-down and bottom-up processing is assumed (see also Burton et al., 1990). All independent processes finally converge into the "cognitive system", which is assumed to play an active role in face recognition by deciding whether an initial match represents a stored familiar face or just an unfamiliar face with a high degree of resemblance to a known face. A range of factors is thought to affect this decision (Young et al., 1985). The model is only concerned about functional components and does not make any inferences about specific localizations of functions in the brain. However, it is influenced by neuropsychological findings of double dissociations between functions.

The model distinguishes between hierarchically organized functional processes and their respective output in form of "codes". The authors propose seven distinct codes that can be derived from faces. The *pictorial code* is generated by any visual pattern and can be understood as the description of a picture. It may contain information about the static pose and expression but also about lighting and overall picture quality. Forced choice recognition tasks on previously unfamiliar faces can be performed on the level of the pictorial code if identical pictures are used. The *structural code* captures a more abstract visual representation of the stable aspects of faces that can be used to distinguish them from one another across a wide range of view-points, head-angles, hairstyles, lighting conditions and other types of pictorial variability. Especially for the recognition of familiar faces the structural code is thought to be essential while the matching of unfamiliar faces has been shown to heavily depend on changeable pictorial cues (Bruce, 1982; Ellis et al., 1979). Bruce and Young (1986) suggest that familiar faces are not represented by one single structural code but a set of interlinked expression independent codes such as distinctive features and global configuration for discrete head angles. Recognition is thought to occur when these encoded structural representations and a set of structural codes match. The *visually derived semantic code* is available both for familiar and unfamiliar faces. It provides information about age, sex, and attractiveness and enables the observer to attribute characteristics such as intelligence, profession, honesty and resemblance to familiar faces. It can be directly influenced by the "cognitive system" that decides which features or components are attended to. The visually derived semantic code is contrasted by an *identity-specific semantic code*, which describes all known details about a familiar person, such as e.g. profession and nationality. The *name code* comprises information about the names of familiar persons. *Expression codes* are available both for familiar and unfamiliar faces and hold information about shapes and postures of internal facial features that underlie emotional expressions. Movements of the lips, tongue and jaw can be used to extract speech information and form the *speech code.* Most importantly, expression and speech codes are proposed to be largely irrelevant for the recognition of faces, because in contrast to the stable structural representations necessary for face recognition they depend mainly on non-rigid facial movements, as represented by changes of internal features over time.

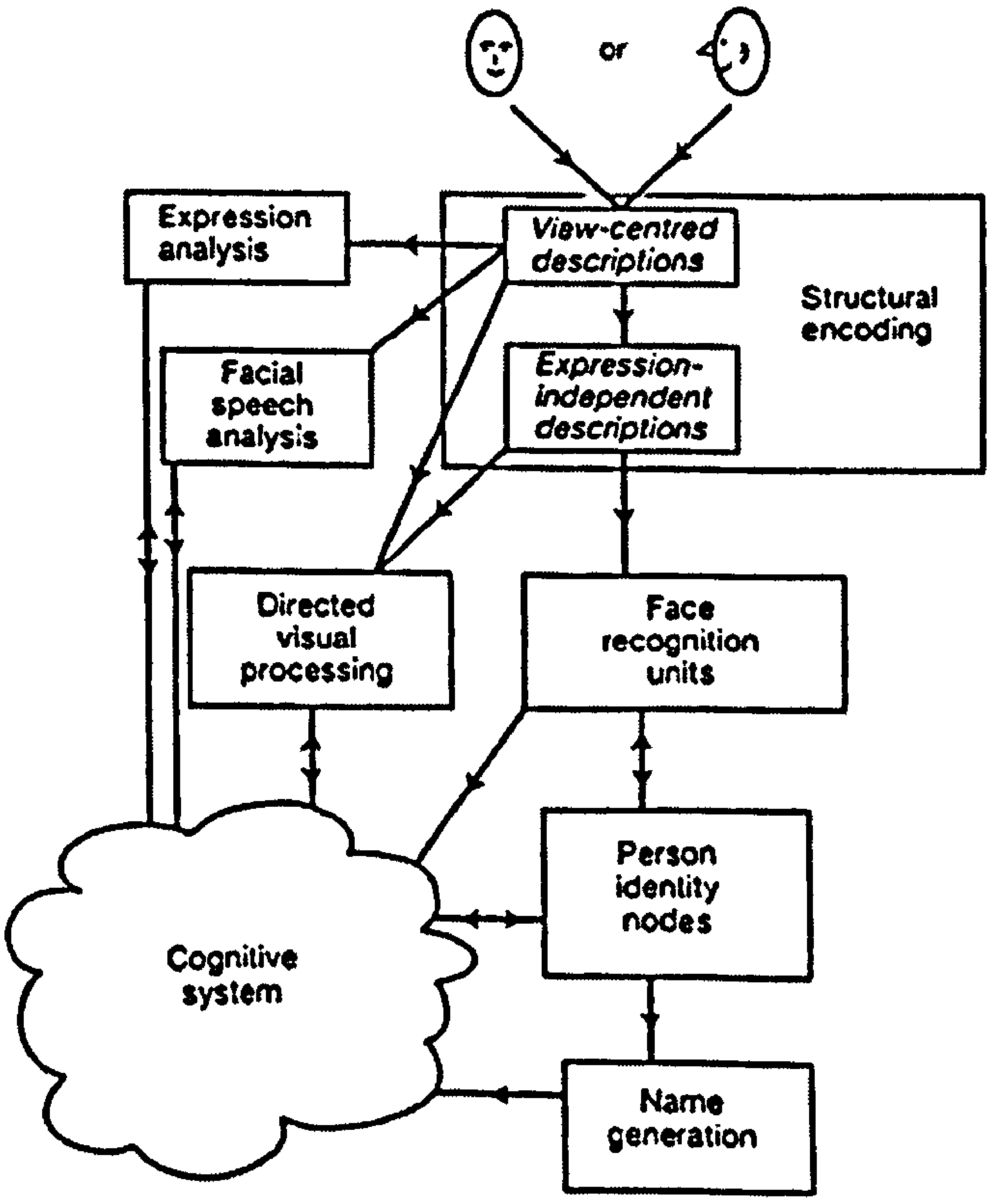

Figure 1: The Bruce and Young (1986) model of face recognition

The model makes the most detailed assumptions about the recognition of familiar faces, which is thought to depend mainly on structural, identity-specific codes. Pictorial, expression and facial speech codes are assumed to play only a minor role for the recognition of identity. At the first stage of processing, a visual encoding of the face is assumed, resulting in “view–centred descriptions” that are used for independent analyses of emotional expression and facial speech.

View centred descriptions undergo further processing and are proposed to become "normalized" with respect to facial expression and facial speech. These “expression independent descriptions” are the basis for directed visual processing on the one hand, but are also crucial for the next stage of the face recognition route on the other. Domain specific and expression independent “Face Recognition Units” (FRUs) are assumed to be composed of expression independent descriptions. It is proposed that there is one expression independent FRU for each familiar face. An FRU can become activated, when a familiar face is seen or when domain general “Person Identity Nodes” (PINs), which contain semantic and biographical information, such as profession, nationality etc. exert top down activation on the FRU. It is further assumed that both FRUs and PINs have a link to the cognitive system, which can directly moderate PIN and indirectly influence FRU activation via the PINs. PINs can be accessed as well from other domains like voices or names (Burton et al., 1990). Finally, in the last stage of the recognition process the name that belongs to a face can be retrieved.

The model and its refined version (Burton et al., 1990) are able to explain a range of empirical findings (Young et al., 1985; for a review see Young, 1998), but some points deserve further investigation and especially the claim of parallel and independent processes has recently been challenged (see section 1.2.6). Also, it is not clear that the identification of a face across different expressions really requires a “normalization” of the view centred descriptions to a stored “expression-independent” Face Recognition Unit. It is possible that idiosyncratic identification features including idiosyncratic expressions may be stored as part of the discriminating identity of an individual at the FRU level.

In the next sections I will try to give a short overview over some of the relevant empirical findings on the processing of facial identity, emotional expressions and facial speech. This must be a rather selective choice, as a huge amount of literature dealing with face perception has accumulated over the last years. Apart from a wide range of behavioural studies, sophisticated imaging techniques have contributed considerably to our understanding of the timing of face perception processes and their underlying neural correlates. Where possible, I will link hypothesized processes (Bruce et al., 1986) to brain structures that might be candidates for the proposed functional modules.

1.2.2 Face recognition

Recognition of familiar faces depends to a large extent on the stable, non-changeable aspects of faces, which is in line with the hypothesis of structural codes as the basis of FRUs. It has been shown that both individual features, sometimes referred to as "first-order" features such as nose, eyes and chin, and "second-order" features, which refer to the arrangement of features, also labelled the facial "configuration", play an important role for face recognition (Cabeza et al., 2000; Carey et al., 1977; Carey, 1992; Hosie et al., 1988; Young, 1987). There is evidence that both for the matching of familiar and unfamiliar faces external and changeable features such as hairstyle are used, but with increasing familiarization, the more stable internal first and second order features gain importance (Ellis et al., 1979; Young et al., 1985). Although faces are three-dimensional structures, recognition of familiar faces from photographs is usually very accurate. Surface information such as pigmentation has been shown to be more relevant than three dimensional shape cues for face recognition (Bruce et al., 1991; Bruce et al., 1994; Liu et al., 2000).

Attempts have been made to investigate neural correlates of the face recognition stages proposed by Bruce and Young (1986). The recent development of imaging techniques such as fMRI has allowed for a precise localization of functions (Gauthier et al., 2001; Kanwisher et al., 1997). Techniques with a very high temporal resolution such as EEG and MEG have provided further insight into the timing of hypothesized processing stages (Eimer, 2000).

Generally, a superiority of the right hemisphere for the processing of facial identity is assumed (Ellis, 1989; Schweinberger et al., 1991). This was often attributed to high visual demands (see e.g. DeRenzi et al., 1968). However there is evidence that also the left hemisphere is involved in face recognition. Prosopagnosia, a rare neurological disorder characterized by selective and severe impairments of face recognition (Bodamer, 1947) is more common after bilateral damage to the inferior temporal cortex (Damasio et al., 1989). However, some cases after damage restricted to the right hemisphere have also been reported (DeRenzi, 1986; Landis et al., 1986; Wada et al., 2001). Possibly, depending on the degree of hemispheric lateralization within an individual, the left hemisphere might be able to compensate damage to face processing areas in the right hemisphere (see also Damasio et al., 1990; DeRenzi et al., 1994; Tovee et al., 1993; Young, 1992). It is widely accepted that the areas for face identity perception are located primarily in infero-temporal brain structures, in particular the gyrus fusiformis, gyrus lingualis and gyrus parahippocampalis (Haxby et al., 1999; Ishai et al., 1999; Kanwisher et al., 1997; Schweinberger, 1992; Sergent et al., 1994). These regions show a higher responsiveness to faces than to other objects such as houses, chairs or tools (Farah et al.,

1999) and it has been suggested that the "fusiform face area" represents a specialized module for face recognition (Kanwisher et al., 1997; McCarthy et al., 1997). It has been shown that the inferior occipital and fusiform gyri are stronger activated in tasks that require the processing of facial identity, while the superior temporal sulcus seems to be more involved in the processing of changeable features of social relevance such as eye gaze (Hoffman et al., 2000). The question whether face recognition forms an innate system of its own, which is independent from the object recognition system, has attracted a lot of interest recently (see also Gauthier et al., 1999; Gauthier et al., 2000a; Gauthier et al., 2000b; Kanwisher et al., 1997; Kanwisher, 2000; Kanwisher, 2001; Tarr et al., 2000).

Differences between face detection and face recognition have been investigated by methods which provide a high temporal resolution, such as event related potentials (ERPs). These methods allow for an exploration of hierarchic processing steps such as the ones hypothesized in the Bruce and Young model (1986). In an ERP study, a face-specific negative potential with a latency of about 200 ms in the left and right fusiform and inferior temporal gyri when recording intracranially from epileptic patients has been reported (Allison et al., 1994). Electrical stimulation of these areas temporarily disrupted the ability to name familiar faces. A slower negative potential around 600 ms post-stimulus onset originated from the ventral occipitotemporal and lateral temporal cortex (Allison et al., 1999; McCarthy et al., 1999; Puce et al., 1999). It is not yet completely clear, whether the relatively face specific N170 deflection (Bentin et al., 1996) originates from the same generator as the intracranial N200 (Allison et al., 1994). It has been demonstrated that the N170 is not dependent on familiarity and probably reflects a correlate of an automatic, attention independent, pre-categorical structural encoding of faces (Eimer, 2000). Processes that differentiate between familiar and unfamiliar faces can be observed at a latency of 250 ms. Immediate repetition of familiar faces generates the N250r, which is more pronounced for repetitions of identical pictures, but is also visible for repetitions of faces on different photographs. It might represent a stimulus triggered access to FRUs (Schweinberger et al., in press) and is also present for different pictures of the same individual. Bentin et al. (2000) recorded an enhanced negativity for familiar compared to unfamiliar faces at a latency of about 400 ms and associated this with the activation of PINs. A similar finding was reported by Eimer (2000) who found an increased negativity between 300 ms and 400 ms and an increased positivity at a latency of beyond 500 ms for familiar compared to unfamiliar faces at midline and parietal electrodes which was also associated with face recognition and identification processes on the FRU and PIN level. Identification and access to semantic information for familiar faces was investigated in an ERP study by Paller et al. (2000). Faces that were learned together with biographical information showed an early posterior and a later anterior positivity when repeated. For faces that were presented without additional information, the effect was restricted to posterior scalp locations. The timing and spatial pattern are consistent with a hierarchical model of face recognition and might point to different neural structures underlying FRUs and Semantic Information Units (Bruce et al., 1986; Burton et al., 1990). The idea of a distinction between modality specific and modality general functional units finds also support from PET (Tempini et al., 1998), ERP (Schweinberger, 1996) and fMRI studies (Shah et al., 2001). There is also fMRI evidence for a functional distinction between the processing of familiar and unfamiliar faces (Leveroni et al., 2000; Shah et al., 2001).

To summarize, it can be said that a range of features trigger face recognition. With increasing familiarity, the representation of stable, internal features and "configurations" becomes more established. Recent imaging studies seem to confirm models of hierarchic processing steps of face recognition. These findings fit quite well with previous reports of face perception difficulties in every day life (Young et al., 1985). There is evidence for different neural substrates for various functional stages. The question whether the face recognition system is independent of a more general object recognition system is currently in the centre of a vivid debate.

1.2.3 Emotional expression

Early scientific reflections on the nature of emotional expressions date back to Sir Charles Bell, who described the role of facial muscles and their anatomy (Bell, 1844). By means of electrical stimulation, Duchenne (1862) demonstrated for the first time that expressions are displayed by contractions of distinctive facial muscles. The first one to propose that expressions are universally recognizable was Charles Darwin. He was aware of the huge importance of the ability to express and interpret emotional expressions for successful social interactions and linked the development of this skill to evolution (Darwin, 1872). However, it took another one hundred years until the universality hypothesis was properly tested (Ekman, 1972) and considerable evidence was found for a largely cultural independence of facial expressions (Ekman, 1982). Today, most researchers agree that there exists a limited set of *basic emotions* (Ekman, 1992; Plutchik, 1980; Tomkins, 1984). In order to distinguish basic emotions from blends of emotions, several criteria have been proposed. Ekman (1992) suggested universality, presence in other primates, distinct physiology, distinct antecedent events, a coherent response pattern, quick onset, brief duration, accompanying distinct appraisal pattern and an unbidden occurrence as necessary criteria (Ekman, 1992). Other authors have added criteria such as distinct ways in which emotions influence perceptions (Izard, 1992; Izard, 1993) or subsequent behavioural action tendencies (Fridja, 1993). On the basis of such criteria, different sets of basic emotions have been proposed, but most vary between a number of five to nine (Lewis, 1993). For studies on face perception, many researchers especially focussed on the six basic emotions happiness, sadness, fear, anger, surprise and disgust, which are displayed in Ekman and Friesen's series of facial affect (Ekman et al., 1976).

The idea of distinct expression categories was not always accepted and a system of two orthogonal emotional axes of facial expressions had also been postulated (Woodworth et al., 1954). According to these authors, facial expressions were located on a pleasant/unpleasant and an attention/rejection continuum, a hypothesis which obviously contradicted the concept of discrete emotional categories. But recently, using the digital picture manipulation technique of "morphing", it has been shown that facial expressions are perceived in a rather categorical manner (Calder et al., 1996), supporting the idea of a limited number of basic emotions.

The relevant facial features giving information on expression seem to differ in some aspects from those relevant for face recognition. Perhaps even more than face recognition, the interpretation of facial expressions depends on combinations of features. While, at least under certain conditions, such as laboratory tasks where a limited set of faces is presented, recognition can be achieved by attending only to a single, distinctive feature, it is hardly possible to interpret an expression just by attending to e.g. the eye region. Pigmentation has proven to be highly informative for the identification of faces (Bruce et al., 1991; Bruce et al., 1994), but it seems to be less important for the interpretation of facial expression (Bruce et al., 1998). Another major difference is that facial expressions are the result of non rigid movement. Nevertheless, static photographs of facial expressions can usually be interpreted with a high accuracy.

With respect to the neural correlates involved in the processing of expression, as for the processing of identity an overall advantage of the right hemisphere is generally assumed (Bowers et al., 1985; Campbell, 1978; Etcoff, 1984; Ley et al., 1979; Natale et al., 1983; Strauss et al., 1981). Some researchers attributed this to the existence of stored representations or "templates" of emotional expressions only in the right hemisphere (Blonder et al., 1991; Bowers et al., 1985; Bowers et al., 1991). However, selective impairments of matching and recognizing emotional expressions have also been reported after unilateral posterior damage in the left hemisphere (Young et al., 1993). Areas that are associated with the ability to perceive facial expressions are the right lateral occipital gyrus and limbic structures, including the amygdala and the basal ganglia (Sergent et al., 1994). A special role of the superior temporal sulcus (STS) for the processing of socially relevant stimuli such as expressions was derived from intracranially recorded ERPs from epileptic patients (Allison et al., 2000) and fMRI experiments (Narumoto et al., 2001; Puce et al., 1998). The STS region might especially interact with the right amygdala (Streit et al., 1999). The idea of discrete basic emotions receives support from studies that suggest different neural substrates for different emotions such as fear, disgust, happiness, sadness and anger, (e.g. Adolphs et al., 1994; Blair et al., 1999; Calder et al., 1996; Morris et al., 1996; Phillips, 1997; Phillips et al., 1998; Sprengelmeyer et al., 1997; Whalen et al., 1998). There is evidence that masked fearful expressions modulate activity in the limbic system, while this seems not to be the case for happy expressions (Whalen et al., 1998). Especially the amygdala has consistently been shown to be involved in fear conditioning both in animals (LeDoux, 1992; Quirk et al., 1997) and humans (LaBar et al., 1998). Disgust seems to be linked to activity in a limbic-cortico-striatal-thalamic circuit and in the anterior insular cortex, an area that is also associated with the processing of smells and visceral stimuli (Phillips, 1997; Phillips et al., 1998). This might reflect the role of disgust for the rejection of potentially unsafe food.

To summarize briefly, the processing of facial expression depends to a large extent on the configuration of facial features. Most researchers agree on a set of about six basic emotional expressions. Possibly, each expression can be associated with activity in at least to some extent distinct neuronal structures that might interact in different ways with structures that process facial identity.

1.2.4 Speechreading

While it is generally known that deaf or hearing-impaired people can learn to use movements of the lips, teeth, jaw and tongue to extract speech information (e.g. Walden et al., 1977), the role of speechreading (or lipreading) in daily conversation is frequently underestimated. Especially under noisy conditions, facial movements may increase speech comprehension to a substantial extent (McLeod et al., 1987; Miller et al., 1955; Sumby et al., 1954). Estimations based on the observed reduction in the minimal signal to noise ratio at which sentences could just be understood when the speaker was visible, compared to the performance when no visual cue was given, suggest a benefit of a considerable 11 dB (Summerfield et al., 1989). Visual speech information may also influence the acoustic perception in incongruent stimulus situations. In the "McGurk illusion", the identification of auditory speech syllables is influenced by the simultaneous presentation of discrepant visible syllables (McGurk et al., 1976). This finding suggests an early and automatic integration of visual and acoustical speech input (see also Calvert et al., 1997; Campbell et al., 1997; Fowler et al., 1991; Green et al., 1991; Massaro, 1987; Meltzoff et al., 1994; Rosenblum et al., 1997; Thompson et al., 1996).

Which kind of information is extracted from faces that can be used to enhance verbal communication? Speech sounds mainly differ with respect to place of articulation (e.g. lips or within mouth), voicing (activity or inactivity of the vocal chords) and manner of articulation (modulation of airflow over time). Manner of articulation and voicing are easier to hear than to see, but for some articulations, place of articulation is easier to see than to hear (Miller et al., 1955), so a multi-modal processing of language enables the speech system to use the maximal amount of information available. It has been stated that humans learn to integrate visual and acoustic speech information very early in life. When given the choice to look at video clips in which visually and acoustically presented sounds either match or mismatch, babies look longer at the congruent stimuli (Kuhl et al., 1982).

One can further differentiate between two relevant types of information for speechreading. *Time-independent features*, which have also been labelled as "static" or "pictorial" are available from photographs and can be used to interpret single vowels or sounds. Early research on speechreading has mainly focussed on static aspects such as lip shape, place of cavity constriction and visible teeth (e.g. Braida, 1991; Massaro et al., 1990; Montgomery et al., 1983; Summerfield et al., 1984). On the other hand, *time-varying features*, also referred to as "kinematic" or "dynamic" reflect the dynamic nature of speech and encompass variations of facial features over time. Although static information can be used in order to classify single vowels and consonants (Campbell et al., 1986; Campbell, 1986; Schweinberger et al., 1998), it is now widely accepted that dynamic aspects play a crucial role for speechreading. It has been shown that speech information can be extracted from point light displays which do not provide static information (Rosenblum et al., 1996a; Rosenblum et al., 1996b). Fusion effects such as the McGurk illusion (McGurk et al., 1976) also occur when pairs of incongruent acoustic and point-light speech stimuli are presented (Rosenblum et al., 1996a). This finding argues for a role of dynamic facial speech information at an early perceptual level. The authors report that static pictures of facial speech do not produce the McGurk effect when paired with acoustically presented consonant-vowel stimuli. This brought them to the conclusion that speechreading from static pictures is a post-perceptual, problem-solving operation (for a detailed review see Rosenblum et al., 1998).

Generally, a specialization of the left hemisphere for speechreading is assumed and there is evidence both from neuropsychological (Campbell et al., 1986; Campbell et al., 1990; Campbell, 1992) and behavioural studies (Burt et al., 1997; Campbell et al., 1996b). It has been shown that non-speech mouth movements activate regions in the superior temporal sulcus (Puce et al., 1998), while speaking faces that are presented without sound additionally increase activity in areas that deal with spoken language, such as the left auditory cortex located in the superior temporal gyrus (Calvert et al., 1997; Campbell et al., 1996b; Ellis, 1989).

To summarize, facial speech is used automatically not only by hearing impaired persons to improve communication. There is evidence for an involvement of speech areas in the left hemisphere and an early and automatic integration of visual and acoustic input. In contrast to face recognition, especially dynamic, time ordered cues provide crucial information for speechreading.

1.2.5 A distributed neural system for face perception

Recently, an attempt has been made to integrate findings on face recognition, facial expression, speechreading, gaze detection and spatially directed attention into a distributed neural system for face perception (Haxby et al., 2000). This system is strongly influenced by the functional model of Bruce and Young (1986) and differentiates between the processing of stable facial features, which underlie face recognition on the one hand, and changeable features that are crucial for the interpretation of expression and visual speech on the other. The authors make an attempt to link these functional aspects of face processing to distinct neural correlates (see also Figure 2). Based mainly on functional brain imaging studies, they put forward the idea of multiple bilateral regions that form a core and an extended system for face processing. According to the authors, the core system consists of three bilateral areas in occipitotemporal visual extrastriate cortex. Each area is assumed to be specialized in different aspects of face perception and to form the neural basis for functional modules as outlined similarly by Bruce and Young (1986).

Haxby et al. (2000) associate the *lateral fusiform gyrus* with the processing of identity and the *superior temporal gyrus* with the representation of changeable aspects of faces such as muscle contractions that are used to express emotions and mouth movements during speech. They hypothesize that the *inferior occipital gyrus* may deal with early processing stages, and due to its anatomical location may provide input to both the lateral fusiform and superior temporal sulcal regions. Additional neural systems, which are not face specific, such as limbic areas and the auditory and parietal cortices are assumed to form the extended system that contributes to the processing of expression and facial speech.

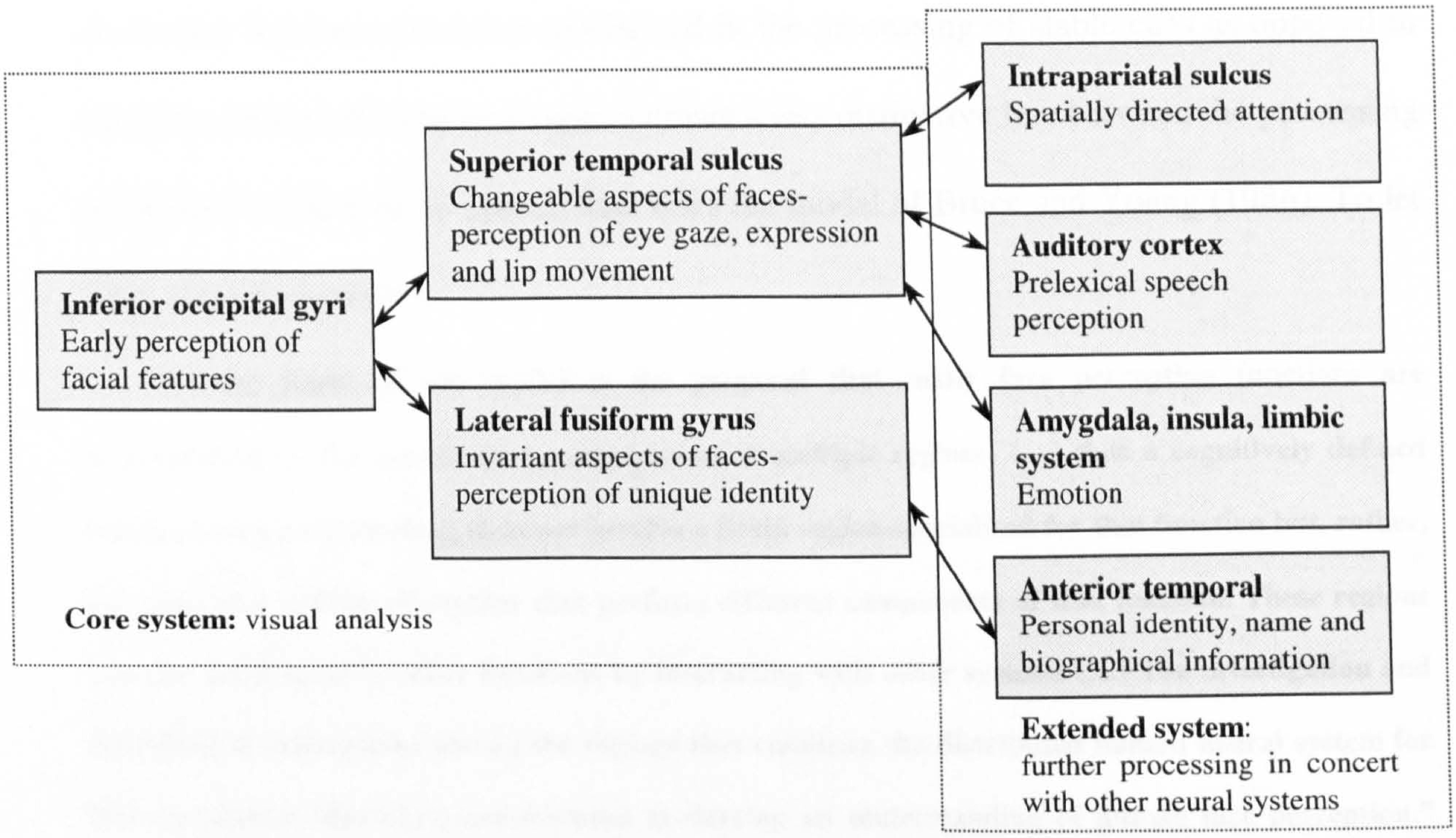

Figure 2: A model of the distributed human neural system for face perception (Haxby et al., 2000). The model consists of a core system, including three regions of occipitotemporal visual extrastriate cortex and an extended system, consisting of regions that are also part of neural systems for other cognitive functions.

Although Haxby et al.'s model (2000) shares some basic aspects with the model proposed by Bruce and Young (1986), such as the idea of an overall hierarchical organisation with distinct functional modules underlying the processing of identity, facial speech and expression, it differs with respect to the importance given to possible interactions between the neural structures associated with the postulated cognitive functions. It leaves open the question to which extent different anatomical substrates are functionally separated from each other. It acknowledges that a face specific region such as the lateral fusiform gyrus might also play a supportive role for the interpretation of emotional expressions, especially for expressions that are strongly associated with a particular individual and that are displayed in a highly idiosyncratic manner. Also, the distinction between structures specialized in the processing of stable cues as opposed to variable and socially relevant aspects draws a less distinctive line between the processing of expressions and facial speech than does the model of Bruce and Young (1986). To let the authors summarize:

> **"At the heart of our model is the proposal that many face perception functions are accomplished by the coordinated participation of multiple regions. (...) thus a cognitively defined function, such as lipreading, does not involve a brain region specialized for that function but, rather, the concerted activity of regions that perform different components of that function. These regions can also participate in other functions by interacting with other systems (...) The investigation and modelling of interactions among the regions that comprise the distributed human neural system for face perception, therefore, are essential to develop an understanding of human face perception."** (Haxby et al., 2000, p. 231).

1.2.6 Evidence pro and contra the independence model

There is both evidence for distinct parallel modules and a partly contingent processing of identity, expression and facial speech. In the following, an overview over some of the most relevant findings from experimental, neuropsychological, physiological and brain imaging studies supporting both sides will be given.

Single cell recordings in the temporal cortex of monkeys have demonstrated specialized cell populations for identity and expression processing although a smaller number of neurons also responded to both kinds of information. Cells that responded primarily to expression were mainly found in the superior temporal sulcus and identity selective cell were mostly situated in the inferior temporal cortex (Hasselmo et al., 1989). Support for the view of a functional independence between speechreading, face recognition and expression processing also comes from neuropsychological studies that show relatively selective impairments and imply that different brain structures are involved in these tasks (Bowers et al., 1985; Bruyer, 1981; Bruyer et al., 1983; Campbell et al., 1986; Etcoff, 1984; Humphreys et al., 1993; Kurucz et al., 1979a; Kurucz et al., 1979b; Parry et al., 1991; Shuttleworth et al., 1982; Young et al., 1993).

In a recent fMRI study a stronger activation of the right superior temporal sulcus was observed when selective attention was directed towards expression compared to identity (Narumoto et al., 2001). Evidence for an independence of identity from expression processing is suggested by a MEG study, which showed that an early midline occipital source distinguished between face and non face stimuli and responded to changes in expressions, while activity in the fusiform area, which is known to be involved in face recognition, was not significantly moderated by varying emotional expressions. (Halgren et al., 2000). Experiments with positron emission tomography (PET) have suggested an activation of different brain regions during the processing of facial identity and emotional expression (Sergent et al., 1994).

Additional support for an independent processing of identity and expression comes from an ERP study, which reports differing effects of immediate stimulus repetitions in identity and expression matching tasks (Potter et al., 1997). While in the identity task an early frontal repetition effect in the latency range of 200-300 ms and a later parietal effect after 350-550 ms was found, only the later effect was present in the expression matching task. In a similar study, topography and timing differed for identity and expression matching (Münte et al., 1998). A negative component was found for mismatches, which had longer latencies in the expression task. The authors interpreted the findings in terms of specialized neural populations underlying identity and expression processing. In a similar experiment, topographical differences between both tasks were found (Bobes et al., 2000), but in contrast to Münte et al. (1998) no timing differences between identity and expression matching were found. The authors attributed this finding to the fact that they had removed external facial features, which selectively made the identity task more difficult. They interpreted the finding of similar RTs in the expression matching task for familiar and unfamiliar faces and an additive effect of familiarity in form of a late positivity and an N400 like component for mismatches in terms of parallel and independent processing of identity and emotional expression.

In summary, there is evidence provided by different methodological approaches both from brain damaged and healthy individuals for the presence of distinct specialized brain areas. However, this is not necessarily a strong argument for an independence on a functional level, as these regions might interact with each other at any processing stage.

Evidence for a functional independence comes from a study by Etcoff (1984), who reported that observers could selectively attend either to facial identity or expression when sorting cards with images of faces into two piles, without much interference from the irrelevant stimulus dimension. Similarly, Young et al. (1986) found faster RTs for familiar compared to unfamiliar faces in an identity-matching task while there was no such advantage for the matching of emotional expressions. Accordingly, Bruce (1986) reported no differences for speeded judgments of emotional expressions from familiar and unfamiliar faces. In a similar study, matching of facial identity was faster for familiar faces, while familiarity did neither improve matching nor classifications of emotional expression and facial speech (Campbell et al., 1996a).

Although these studies appear to support the hypothesis of a functional independence of the perception of facial identity, emotional expression and facial speech, some recent data seriously challenge this view. It has been shown that the McGurk illusion (McGurk et al., 1976) was significantly reduced for familiar faces when the voice of a different person was presented (Walker et al., 1995). The authors interpreted these results as evidence for an influence of facial identity on speechreading. The result is of particular interest, because cross-gender combinations do not reduce the McGurk illusion when both face and voice are unfamiliar (Green et al., 1991). A recent study suggests that characteristic facial movements can be used to identify a speaker, even when shape and texture information is eliminated using point-light displays (Rosenblum et al., 2002). Moreover, a better speechreading accuracy was observed when speaker identity was held constant in contrast to trial-to-trial speaker variations (Yakel et al., 2000), which might argue for an early integration of speaker identity and facial speech.

It has also been demonstrated that performing an identity classification task on hybrid face stimuli (gender information is given e.g. in the low and expression in the high spatial frequency range) can influence the preferred spatial frequency in a following expression classification task (Schyns et al., 1999). In a recent study, in which pair wise presented faces were matched either for identity or emotional expression, reaction time was not independent of the respective irrelevant dimension (White, 2001). For "same" responses, RTs were faster when also the task-irrelevant feature was the same while "different" responses were made faster when both relevant and irrelevant features differed. Unfortunately, the design of that study does not allow us to rule out that the results might have been due to the use of superficial pictorial cues. A superior speechreading performance when subjects were personally familiar with talkers was observed and interpreted in terms of a parallel-contingent processing of identity and facial speech (Schweinberger et al., 1998). In the same study, a slowing of speechreading and expression classifications when face identity was varied in comparison to a control condition where identity was held constant was observed. However, varying emotional expressions or facial speech movements did not influence identity judgments. The authors concluded that identity might be perceived independently of, but exerts influence on expression analysis and speechreading.

To summarize, the view of a clear cut functional independence of facial identity processing, speechreading and the analysis of facial expression has recently been challenged by studies suggesting contingencies between these dimensions. There is some accumulating evidence that at least the output of identity related processes may influence the perception of expression and facial speech. At the time Bruce and Young (1986) suggested the independent parallel model of face recognition, concepts of higher visual processes tended to describe processing stages as strictly serial, separate, and independent. In contrast, more recent theoretical approaches put forward cascade models, for which a complete categorization is not required to proceed to a later stage. They also underline that interactive systems can allow for a range of cross-talk between higher and lower levels of processing. There are now some hints into the direction of such cross talks between facial identity processes on the one hand and expression and facial speech processes on the other.

1.3 Rationale for testing the independence model

The parallel model of face perception (Bruce and Young, 1986) makes a range of testable predictions on which the following experiments are based. First, familiarity should have an influence on RTs in face recognition tasks, because the model assumes that for each known face there is a FRU signalling familiarity in a fast and automatic manner, while more time consuming directed visual processes are involved in rejecting a face as unfamiliar. No influence of familiarity is expected on RTs in expression and speechreading tasks because these processes are thought to be independent from FRU and PIN output. Accordingly, face recognition should be independent of displayed emotional expression, because stored modality specific representations in form of FRUs are supposed to be composed of the structural, expression independent codes and a "normalization" of the visual input with respect to emotional expressions is assumed. Therefore, it will be explored, whether face familiarity interacts with either speechreading or the processing of facial expression.

A different approach to distinguish between parallel or integrated processing of functional components is to test, whether task-irrelevant stimulus dimensions can influence the processing of selectively attended dimensions. A strict parallel model predicts that task-irrelevant variations of one facial dimension such as identity, expression or facial speech, do not interfere with the processing of another task-relevant dimension. If selective attention cannot be directed towards one particular dimension, this would argue for an at least partly integrated processing (see also Garner, 1974; 1976). To test this prediction, the selective attention paradigm, which is described in more detail in section 1.3.1 will be used. There is recent evidence that both speechreading

(Schweinberger et al., 1998; Yakel et al., 2000) and the processing of facial expression (Schweinberger et al., 1998; Schweinberger et al., 1999) might to some extent be contingent on face identity processing.

Dynamic information is assumed to play a major role for speechreading (for a review see Rosenblum et al., 1998). Many studies on speechreading however have used static pictures as stimuli. The ecological value of such pictures for the processing of facial speech is therefore questionable. To my knowledge, no reaction time experiments on an influence of facial identity on speechreading speed using dynamic clips have been conducted so far. It might be possible that facial speech units interact with FRUs, when the appropriate dynamic input is provided, so I am going to close this empirical gap.

Finally, it can be assumed that manipulations of one facial dimension such as expression, should leave classifications of other dimensions such as identity unaffected. This prediction can be tested by using the „morphing“ technique, which allows for a manipulation of selected facial dimensions. The technique is outlined in more detail in section 1.3.2.

With respect to the model of an independent processing of facial identity, emotional expression and facial speech (Bruce and Young, 1986), the following predictions will be tested:

1. Speechreading speed is uninfluenced by personal familiarity.
2. Task-irrelevant speaker variations do not influence speechreading speed both for familiar and unfamiliar faces.
3. Classification response times and classifications of emotional expressions do not differ between familiar and unfamiliar faces.

4. Classification response times and classifications of facial identity do not differ for various expressions.
5. Task-irrelevant identity variations do not interfere with classifications of emotional expression.
6. Task-irrelevant variations of emotional expression do not interfere with classifications of facial identity.

1.3.1 The selective attention paradigm

Predictions 2, 5 and 6 will be tested by applying Garner's selective attention paradigm (Garner, 1974; Garner, 1976). In this paradigm participants are required to make speeded two-choice classifications of four types of stimuli representing the crossing of two different dimensions. In the classic version of the paradigm, the stimuli are presented in three different experimental conditions. In the *control condition*, the stimuli vary along only the respective relevant dimension, while the irrelevant dimension is held constant. Applied to face perception, for example only pictures of Person A displaying two expressions might be shown in the control condition. In the *orthogonal condition* stimulus sets are presented that include variety with respect both to the relevant and the irrelevant dimensions (e.g. pictures of two individuals displaying two expressions). In the *correlated condition*, there is a co-variation between the two dimensions; for example, within a block of trials there are only pictures of Person A displaying a happy expression and of Person B displaying a neutral expression. The point of interest in this paradigm is, how well participants are able to process the relevant dimension independently of variations in the irrelevant one. With respect to the comparison between the control and the correlated conditions, a so-called redundancy gain (e.g. faster RTs in the correlated condition) is usually considered as an indication of integrated processing, suggesting that at some level the combination of features is perceived as a unitary event (e.g. Etcoff, 1984; Garner, 1974). However, it has been pointed out that a redundancy gain might also occur in the context of parallel and independent processing of both dimensions. If perceivers classify the stimuli in the correlated condition by systematically using the faster or more discriminative dimension for each trial (in spite of the task instruction), a redundancy gain might also be observed (Eimas et al., 1978; Green et al., 1991). Consequently, a redundancy gain is consistent with, but should not in general be considered a strong indication of integral processing. A more important indicator of an influence of the irrelevant dimension on the processing of the relevant one is the comparison of the control and the orthogonal condition. An increase in reaction times for orthogonal compared to control trials shows that variation along the irrelevant dimension influences the classification of the relevant dimension. In other words, selective attention to the relevant dimension is impossible; presumably indicating that both dimensions are processed in an integral manner (Garner, 1976; Green et al., 1991). In contrast, similar RTs for the control and orthogonal condition indicate that variation along the irrelevant dimension does not interfere with the perception of the relevant dimension. In this case, the two perceptual dimensions are assumed to be processed separately.

1.3.2 Morphing

Morphing provides a different approach for the testing of an independence of stimulus dimensions as it allows for a selective manipulation of facial dimensions in realistic stimuli. The technique makes it possible to create face stimuli with a controlled perceptual saliency of a particular facial dimension such as e.g. expression.

Morphing can create a photographic-quality continuum between any two images. The morphing procedure has two components, *warping* and *fading*. Warping basically involves a spatial transformation of control points from their original position in one image to their final position in the other image (see also Figure 3). Fading refers to a linear transition of all corresponding pixel values between start and end image (for details, see e.g. Beale et al., 1995).

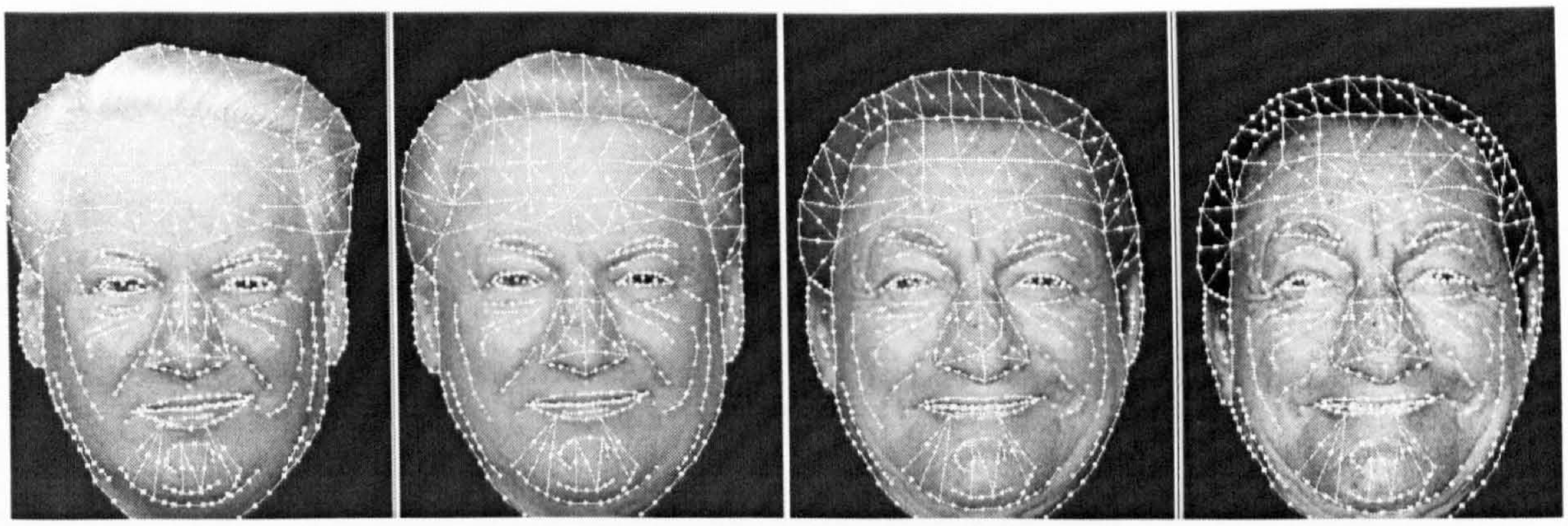

Figure 3: Examples of start, intermediate and end pictures (100:0%, 71:29%, 29:71% and 0:100%, from left to right) of an identity morph continuum, including control points and lines. Intermediate morph stimuli are produced by *warping* between control points and *fading* between all corresponding pixel values.

A typical observation is that most images on a continuum between two identities are consistently categorized as belonging to one of the two people corresponding to the original stimuli at each end of the continuum. There is only a relatively narrow area of ambiguity with respect to facial identity (Beale et al., 1995; Schweinberger et al., 1999) and similar observations have been reported for the perception of emotions in morphed faces (Calder et al., 1996; Etcoff et al., 1992; Schweinberger et al., 1999; Young et al., 1997). However, Young et al. (1997) and Schweinberger et al. (1999) also showed that RTs slowed down with increased distance from the endpoints of the morph continuum, even for the faces that were still consistently classified as belonging to one category, reflecting a selectively decreased perceptual salience of these intermediate morph stimuli. This effect will be used in Experiments 11 and 12 in order to investigate influences of perceptual saliency and relative processing speed in the Garner paradigm.

2 Experiments 1-3: Influences of speaker variations on speechreading speed. Experiments with moving faces.

2.1 Purpose of Experiments 1 to 3

Recent studies have suggested an influence of task-irrelevant talker variability on speechreading (Schweinberger et al., 1998; Yakel et al., 2000). The aim of the following experiments was to clarify the influence of task-irrelevant speaker variations on speechreading speed using a similar design as the experiments mentioned above. However, important limitations of the cited studies were that no reaction time measurements were taken (Yakel et al., 2000) or that only static pictures were presented (Schweinberger et al., 1998). Although some information about facial speech can be extracted from photographs, the nature of speech is dynamic and one may miss the most relevant information if only static displays are used (Rosenblum et al., 1998). It is therefore of major interest to see whether the reported findings generalize to dynamic stimuli. In order to test influences of task-irrelevant speaker variations on speechreading speed for moving faces, software was developed to measure reaction times to video clips with millisecond accuracy.

Another concern with the study by Schweinberger et al. (1998) is the small set of only eight face stimuli per experiment, which may have encouraged picture based response strategies. Experiments 1 to 3 therefore used larger sets of 48 stimuli per experiment.

2.2 Rationale of Experiments 1 to 3

Experiments 1 to 3 investigated the influence of task-irrelevant speaker variations on speechreading performance. Video digitised faces were presented either in *static*, *static-sequential* or *dynamic* mode and participants performed speeded classifications on vowel utterances as representing /u/ or /i/[1]. A Garner type interference paradigm was used in which speaker identity was task-irrelevant but could be correlated, constant, or orthogonal to the vowel uttered.

In each of the experiments 48 different stimuli were used in order to discourage picture based response strategies. Experiment 1 was designed to determine whether the reported influence of identity on the perception of facial speech in static pictures (Schweinberger et al., 1998) is also found when picture based strategies are discouraged by increasing the overall stimulus set. Experiment 2 addressed the question of whether the effects found with static pictures can be generalized to more realistic moving stimuli. While identity can usually be easily derived from static faces, the additional information gained from non-rigid movements (dynamic changes within the face such as mouth movements) is especially important for speechreading (for a review see Rosenblum et al., 1998). As the nature of facial speech is dynamic, moving faces represent more realistic stimuli than static material. Another potential difference between static and dynamic stimuli is that, while in static pictures identity and facial speech information are available at the same time, speaker identity may be available prior to the onset of facial speech information in dynamic stimuli. An additional experiment was therefore needed in order

---

[1] In German pronunciation

clearly to determine the influence of movement itself on speechreading performance. In Experiment 3, *static-sequential* stimuli were presented, which consisted of two static images, with the first image showing the face before the articulation and the second image showing the face at the apex of an utterance. Therefore, just as in the dynamic clips, facial identity information was available before the onset of facial speech information.

As mentioned, in static pictures such as the ones used in Experiment 1, information about identity and facial speech is available at the same time, and this difference between static and dynamic stimuli allows testing for a number of alternative interpretations of Schweinberger et al.'s findings (1998). There is some evidence that identity might be processed faster than facial speech (Campbell et al., 1996a), perhaps because identity information might be processed in a highly automatic manner. If identity and facial speech were processed in a serial manner and the processing of identity was mandatory and faster, an asymmetric interaction as described by Schweinberger et al. (1998) would be expected. In this case, task-irrelevant trial-to-trial variation of identity would increase the processing time of this stage and influence all later processes, even if these were independent from the output of identity modules. Task-irrelevant variations of facial speech would have no influence on identity classifications, because identity processing would already be completed before facial speech information is accessed. Although a serial processing of face identity and facial speech does not seem to be a likely scenario (see also chapter 1), this possibility was examined in Experiments 2 and 3, where facial speech information was given one second after face onset. In this setting, the processing of identity should already be finished when the speech classification has to be made. If the orthogonal interference of identity on speechreading found with static portraits

(Schweinberger et al., 1998) is due to serial processing of identity and speechreading, it should therefore disappear both in dynamic and static-sequential presentation mode.

Another possible interpretation for the results of Schweinberger et al. (1998) would be a short-term perceptual tuning to speaker-specific speaker characteristics. For the acoustic modality it has been shown that recognition of phonemic properties becomes more difficult in blocks that contain trial-to-trial speaker variations. However, no decrease in performance has been found if there was a variety of speakers within the block, but the speaker was held constant over a number of trials (Green et al., 1997). It was argued that this effect is due to the early encoding of individual speaker characteristics, which are held in working memory in order to facilitate the encoding of acoustic properties and their conversion into phonemic codes. If the speaker is held constant across trials, it is possible to make use of the speaker-characteristic representations still active in working memory. In contrast, trial-to-trial speaker variation may result in delays in phonetic processing, because characteristic properties have to be encoded for every trial anew. If such a perceptual tuning also takes place in speechreading and indeed accounts for the orthogonal interference, no such effects should be found for dynamic and static-sequential faces if identity information is given before facial speech onset. Perceptual tuning could also occur in such multiple speaker lists, because extra time to process identity specific characteristics is provided. While both dynamic and static-sequential conditions control for serial processing and perceptual tuning to identity-specific speaker characteristics, the comparison between these conditions would allow for a clearer investigation of the influence of movement on orthogonal interferences.

In sum, four potential patterns of results would yield the most straightforward interpretations. If the influence of identity variations on speechreading were the same in all three presentation modes, the data would argue for a generalization of the influence of irrelevant speaker variations on speechreading (Schweinberger et al., 1998) over a variety of stimulus situations. Should the influence of identity variations disappear if identity information is given prior to facial speech information in dynamic and static-sequential presentation modes, this would argue for the influence of perceptual tuning or a serial processing of facial identity and speechreading. If there is no orthogonal interference in the static condition with the larger number of 48 stimuli, this would suggest that the results of Schweinberger et al. (1998) might have been mainly influenced by picture based response strategies.

Finally, a difference between the influence of identity variations on speechreading in static-sequential and dynamic presentation mode would mean that the dynamic information in itself modifies the interaction between the processing of identity and facial speech.

2.3 Methods

2.3.1 Participants

Eighteen participants (seven women and eleven men) aged 21-33 years (*M* = 23.7, *SD* = 3.0 years) contributed data in Experiment 1. Eighteen different participants (fifteen women and three men) aged 20-36 years (*M* = 24.9, *SD* = 5.0 years) took part in Experiment 2. Another eighteen subjects (eight women and ten men) aged 20-46 years (*M* = 26.7, *SD* = 7.9 years) contributed data in Experiment 3. Participants were randomly assigned to Experiment 1, 2, or 3 and received either a fee of ten deutsche marks (DM; *n*

= 40) or course credit (*n* = 14). All experiments were conducted at the University of Konstanz, Germany, and all participants were native speakers of German.

In Experiment 2, data from two additional participants had been replaced due to an excessive rate of outliers (in some experimental conditions more than 20% of all trials outside the range of 150 to 1500 ms, as compared to an average of 0.4%). In Experiment 3, two additional subjects had been replaced because of excessively slow RTs (RTs exceeded mean RTs for more than two standard deviations).

2.3.2 Stimuli and Apparatus

The stimuli consisted of digitised video clips of faces of two young female volunteers. The clips were directly recorded with a capture rate of 25 frames/sec on hard disc using an *AV Master*™ video capture card. For both volunteers, clips of two vowel articulations (German vowels /u/ and /i/) were recorded. For the raw videos, speakers were instructed to produce speech with a consistent timing, producing no speech movements for an initial two seconds period after which they articulated a vowel. For each vowel and volunteer, twelve different video clips were prepared. One video clip each was taken for three viewpoints (frontal, ¾ left, and ¾ right profile view), two hair covers (with and without hat), and two versions of eye gaze (looking directly into the camera and looking to the side), resulting in a set of 48 video clips (twelve versions x two vowels x two speakers).

The original clips were digitally edited using the *Ulead Media Studio*™ Software. In all experiments the edited stimuli consisted of video clips of 3000 ms (75 frames at a rate of 25 frames per second). In all trials a white fixation cross on a black background was shown for one second, followed by a face stimulus visible for 2000 ms. Faces were shown in a stimulus area of 12.2 by 9.2 cm at a screen resolution of 800 by 600 pixels. At a viewing distance of about 60 cm this corresponded to a horizontal visual angle of 11.5 degrees and a vertical visual angle of 8.7 degrees.

For the purpose of the experiments, clips were presented by *MS-DOS*™ based video software (*QuickView*™) on an IBM compatible personal computer (PC1). This presentation computer was connected to a second PC (PC2) that controlled the experiment and measured reaction times with millisecond accuracy. A trigger synchronized to the vertical retrace of the PC1 presentation monitor was sent to PC2 immediately before presenting the first video frame, initiating RT measurements.

In Experiment 1 (static presentation), clips consisted of a fixation cross that was presented for 1000 ms, followed by a static face for 2000 ms. That face was the frame from the original clips which showed the apex of the vowel utterance. The preceding 1000 ms in which the fixation cross was presented were subtracted offline so that reaction time was adjusted to the onset of the articulating face. Examples of the static pictures used in Experiment 1 can be seen in Figure 4.

Figure 4: Examples of stimulus material used in Experiment 1 (static presentation). Top row, columns from left to right: Portraits of Speaker A, uttering /i/, /u/, /i/, /u/. Bottom row, columns from left to right, Portraits of Speaker B, uttering /i/, /u/, /i/, /u/. Original stimuli were in colour.

In Experiment 2 (dynamic presentation), clips consisted of a fixation cross that was presented for 1000 ms, followed by a complete sequence of face movement during the articulation of /u/ or /i/. At face onset, a static face was shown for the first 1000 ms, followed by 1000 ms of dynamic facial speech. Within the clip, the first frame of the sequence on which the mouth started to open was always presented at 1000 ms after face onset. The preceding 2000 ms (in which the fixation cross and the static face were presented) were subtracted offline so that reaction time measurement was adjusted to articulation onset. In order to obtain comparable temporal characteristics for all clips, the average time from movement onset to the apex across all raw video clips taken from both speakers was calculated. Where necessary, individual clips were then edited in order to synchronize the apex of the articulation to the empirically found average value of 320 ms (time from mouth opening to apex).

For Experiment 3 (static-sequential presentation), two pictures were extracted from each of the described 48 dynamic clips. The first picture showed the first face frame of the corresponding dynamic clip (closed mouth), and the second picture showed the frame corresponding to the articulation apex. After a fixation cross that was present for 1000 ms, both pictures were presented sequentially for 1000 ms each and reaction time was adjusted offline to the onset of the articulating face by subtracting 2000 ms from the total RT. The timing properties of static, static-sequential and dynamic clips are illustrated in Figure 5.

2.3.3 Design and procedure

The experiments reported in this article were designed according to the selective attention paradigm reported by Garner (1974, 1976; see section 1.3.1). In the present experiments, the stimuli were faces that varied along the dimensions of speaker (Speaker

A or Speaker B) and vowel (/u/or /i/). In all experiments, participants were instructed to classify the faces with respect to vowel articulation by pressing one key for /u/ and another key for /i/ utterances. Half of the participants pressed the left "Ctrl" key of the keyboard for /i/ and the right "Ctrl" key for /u/ utterances while for the other half this assignment was reversed. Subjects were informed that two different speakers would be seen but that they should selectively attend to vowel articulation while disregarding speaker identity.

Stimuli were presented in three experimental conditions. Each condition consisted of two blocks of 96 experimental trials and all 24 clips per speaker were presented four times per condition. (There were ten additional catch-trials per block in the correlated condition, as described below). Each block started with an additional fifteen practice trials which were randomly selected from the stimulus pool and which were not analysed.

In the *control condition* one block showed only utterances of Speaker A and the other block consisted of utterances of Speaker B. In the *correlated condition,* there was a covariance between speaker and vowel: in one block, /u/ utterances were made by Speaker A and /i/ utterances by Speaker B, while this pattern was reversed in the second block. In order to control for strategies, which, against task instructions, would use speaker identity as relevant criteria, an additional ten trials per block violated the rule of covariance between both stimulus dimensions and acted as *catch-trials.* A strategy of systematically using the irrelevant dimension would show up in large performance costs in these catch-trials. The *orthogonal condition* consisted of two identical blocks in which both speaker and vowel were varied orthogonally. This procedure made sure that for each condition exactly the same stimuli were used, ruling out any potential confound with stimulus differences. The order of blocks within experimental conditions was randomly varied across participants, and the order of experimental conditions was completely counterbalanced across participants. Breaks were allowed after each block.

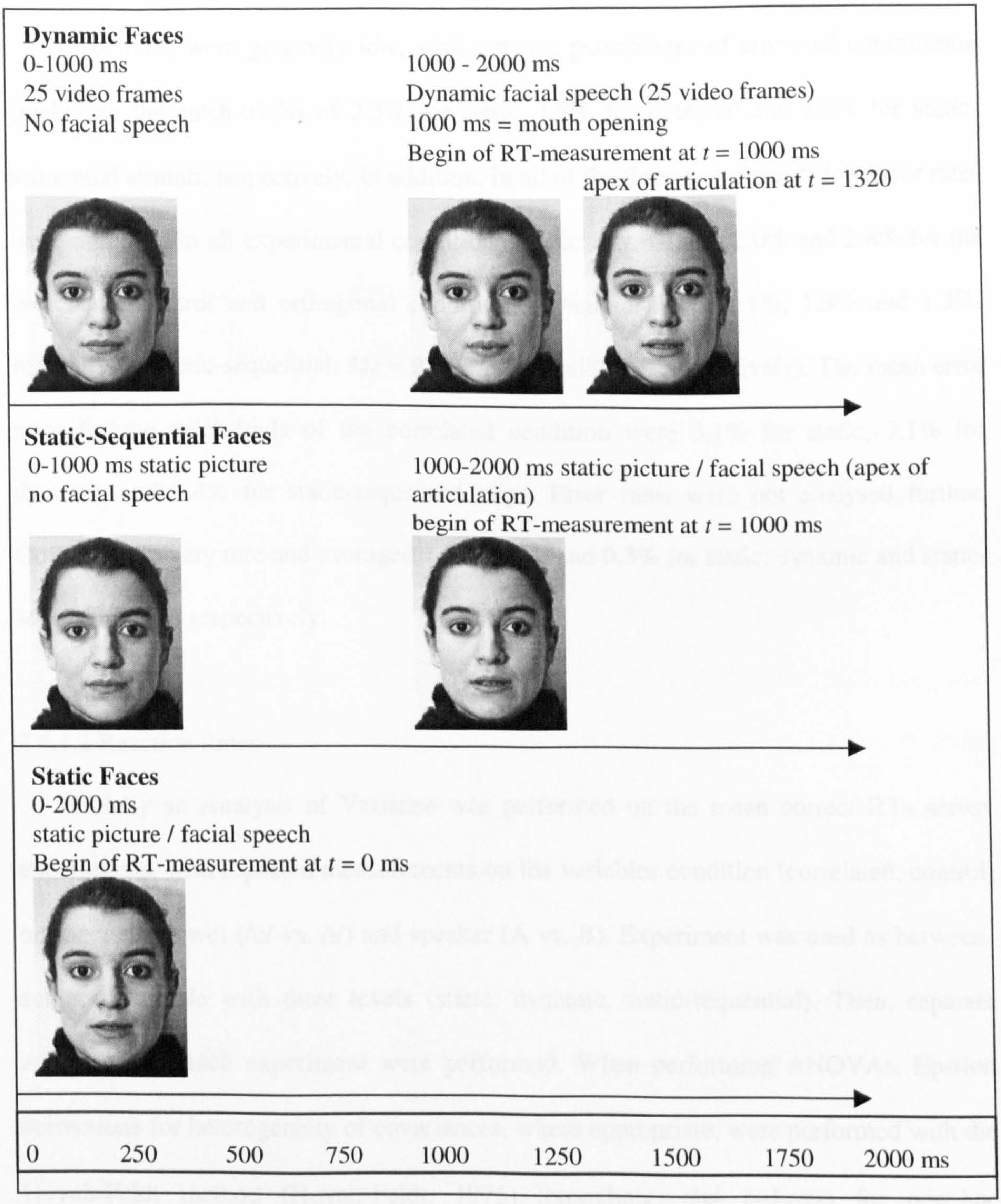

Figure 5: Schematic illustration of timing properties for dynamic, static-sequential and static presentation mode.

2.4 Results

2.4.1 Effects of condition

2.4.1.1 Error rates

Error rates were generally low, with average percentages of errors of commission (including the catch-trials) of 2.3% for static, 1.8% for dynamic and 2.8% for static-sequential stimuli, respectively. In addition, in all of the three experiments low error rates were observed in all experimental conditions (static: *Ms* = 1.4%, 2.0% and 2.8% for the correlated, control and orthogonal conditions; dynamic: *Ms* = 1.1%, 1.9% and 1.3%, respectively; static-sequential: *Ms* = 2.1%, 2.2% and 2.6%, respectively). The mean error rates for the catch-trials of the correlated condition were 3.1% for static, 3.1% for dynamic and 4.4% for static-sequential clips. Error rates were not analysed further. Outliers were very rare and averaged 1.3%, 0.4% and 0.3% for static, dynamic and static-sequential clips respectively.

2.4.1.2 Reaction times

Initially an Analysis of Variance was performed on the mean correct RTs across experiments, with repeated measurements on the variables condition (correlated, control, orthogonal), vowel (/u/ vs. /i/) and speaker (A vs. B). Experiment was used as between-subjects variable with three levels (static, dynamic, static-sequential). Then, separate ANOVAs for each experiment were performed. When performing ANOVAs, Epsilon corrections for heterogeneity of covariances, where appropriate, were performed with the Huynh-Feldt method (Huynh-Feldt, 1976) throughout, and α–levels for post-hoc ANOVAs were Bonferroni corrected.

The mean RTs of all conditions for Experiments 1 to 3 are plotted in Figure 6. Although the mean condition effect is of major interest, the data are also plotted for every combination of speaker and vowel in order to show the variability of the condition effect. The overall ANOVA revealed a main effect of experiment, $F(2, 51) = 30.7, p < 0.001$. Post-hoc analysis by Duncan's multiple range test ($\alpha = 0.05$) indicated that RTs for static stimuli (675 ms) were longer than RTs both for dynamic (491 ms) and static-sequential stimuli (474 ms), which did not differ significantly from each other. Overall, there was a strong condition effect, $F(2, 102) = 12.5, p < 0.001$. Post-hoc analysis using Duncan's multiple range test ($\alpha = 0.05$) revealed that this main effect was due to the fact that RTs for the correlated condition were reliably faster (531 ms) than those for the control condition (546 ms), which were significantly faster than those for the orthogonal condition (562 ms). The condition effect did not interact significantly with experiment, $F(4, 102) < 1$.

There was a main effect of speaker, $F(1, 51) = 81.0, p < 0.001$, indicating that participants showed somewhat faster RTs for Speaker A as compared to Speaker B (*Mdiff* = 24 ms). This was qualified by a significant interaction between experiment and speaker, $F(2, 51) = 27.0, p < 0.001$. Speaker differences were particularly clear in the dynamic presentation mode, *Mdiff* = 51 ms, compared to static (*Mdiff* = 15 ms) and static-sequential stimuli (*Mdiff* = 5 ms). There was also a significant interaction between experiment and vowel, $F(2, 51) = 10.1, p < 0.001$. Accordingly, it was more difficult to recognize /i/ compared to /u/ vowels when presented dynamically (*Mdiff* = 33 ms) compared to static-sequential (*Mdiff* = 3 ms) and static presentation (*Mdiff* = 13 ms).

Overall, there was a strong interaction between vowel and speaker, $F(1, 51) = 174.5$, $p < 0.001$. For Speaker A, participants were faster when classifying /i/ compared to /u/

utterances (520 ms vs. 549 ms), while for Speaker B this effect was reversed (581 ms for /i/ vs. 537 ms for /u/ utterances). This interaction was moderated by a three-way interaction between condition, vowel and speaker, $F(2, 102) = 6.5$, $p < 0.01$. Although the interaction was present in all conditions, it seemed to be less pronounced in the control condition (see also Figure 7). Finally, the interaction between vowel and speaker was further qualified by a significant three-way interaction between experiment, vowel and speaker, $F(2, 51) = 8$, $p < 0.001$. No other effects or interactions were significant.

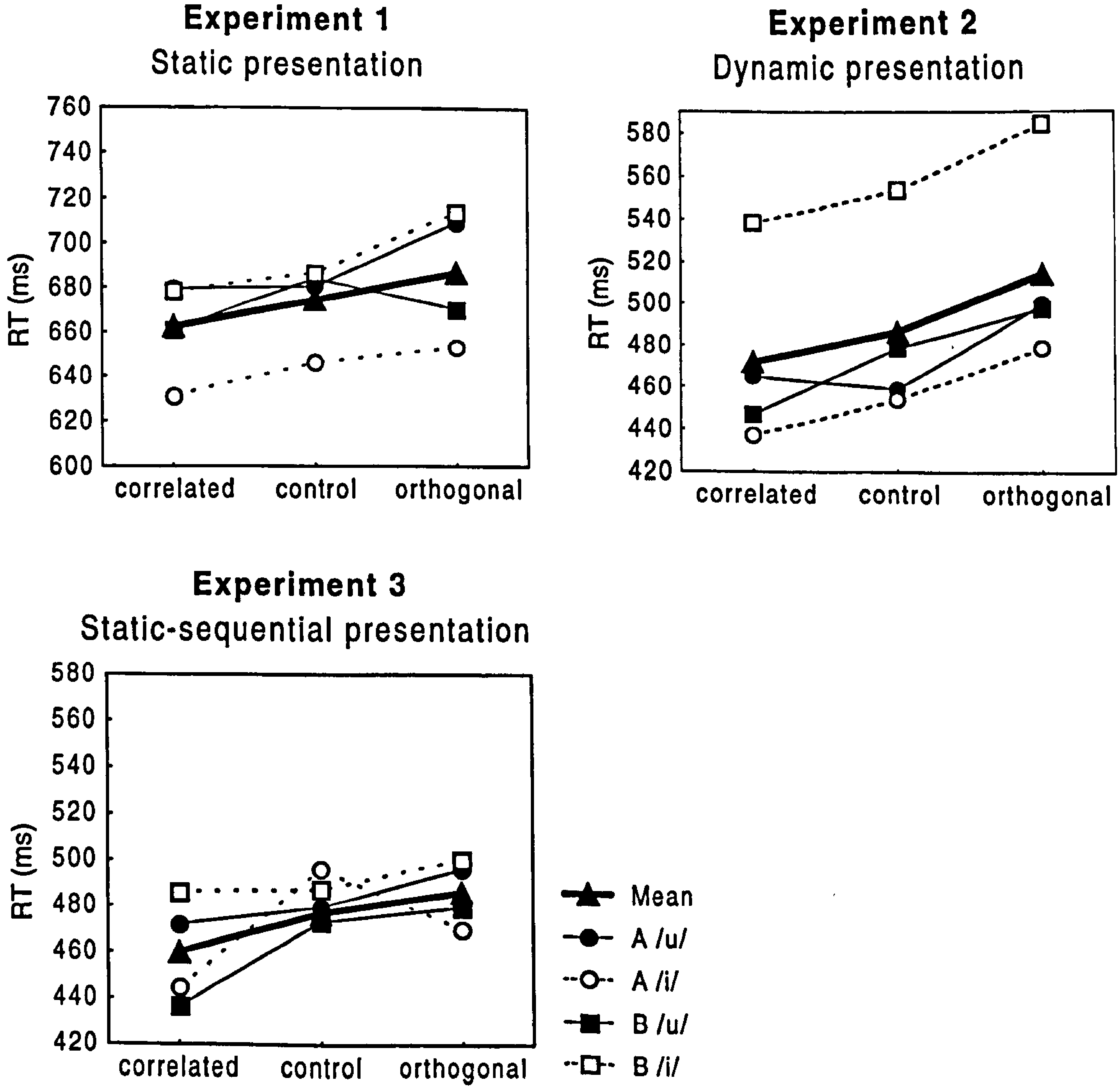

Figure 6: Mean reaction times (RTs) in Experiments 1-3 for the experimental conditions and every combination of vowel (/u/ or /i/) and identity (Speaker A or B). Bold lines show means across all combinations of both dimensions. Note that y-axis differs for Experiment 1.

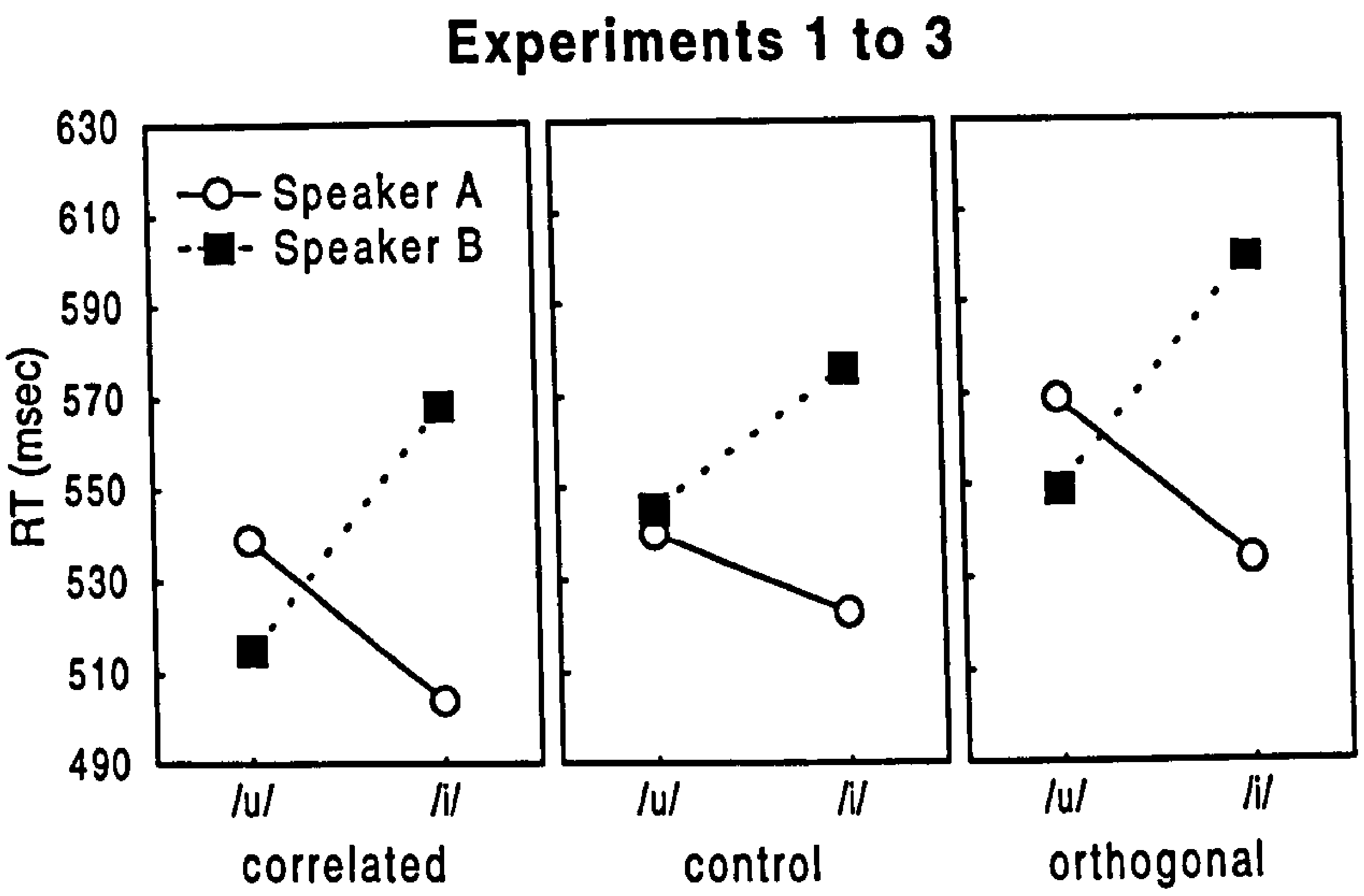

Figure 7: Interaction between condition, speaker and vowel for RTs, averaged across Experiments 1 to 3. Note that vowel-specific differences in RTs for both speakers are smallest in the control condition.

In the following, results for separate ANOVAS for each experiment will be reported. Repeated measurement variables were condition (correlated vs. control vs. orthogonal), vowel (/u/ vs. /i/) and speaker (Speaker A vs. Speaker B).

2.4.1.2.1 Static presentation (Experiment 1)

There was a numerical increase in RTs from the correlated (662 ms) to the control (674 ms) and the orthogonal condition (687 ms). Although the effect was only marginally significant, $F(2, 34) = 2.8, p < 0.08$, it was further explored by Duncan's Multiple Range post hoc tests ($\alpha = 0.05$). These tests suggested a significant difference between the correlated and the orthogonal, but not between the control and the orthogonal condition. Overall, RTs where faster for Speaker A, $F(1, 17) = 11.1, p < 0.01$ (667 ms vs. 682 ms for Speaker A and B, respectively). The speaker main effect was further modulated by a significant two-way interaction between speaker and vowel, $F(1, 17) = 30.3, p < 0.001$. Reaction times were shorter for Speaker A when pronouncing /i/ compared to /u/ (644 ms vs. 690 ms), while for Speaker B the opposite was the case (693 ms vs. 672 ms, for /i/ and /u/, respectively). There was a significant three way interaction between condition, speaker and vowel $F(2, 34) = 3.4$, $p < 0.05$. No other effect was significant.

2.4.1.2.2 Dynamic presentation (Experiment 2)

There was a main effect of condition, $F(2, 34) = 6.38$, $p < 0.01$. A post-hoc analysis using Duncan's Multiple Range test ($\alpha = 0.05$) attributed this effect to shorter RTs both for the correlated (472 ms) and the control condition (487 ms) compared to the orthogonal condition (515 ms). The vowel main effect, $F(1, 17) = 40.1$, $p < 0.001$, indicated that /u/ utterances were classified faster than /i/ utterances (*Mdiff* = 33 ms). There was also a significant main effect of speaker, $F(1,17) = 76.9$, $p < 0.001$, showing overall slower RTs for utterances of Speaker B (*Mdiff* = 51 ms). The two-way interaction, $F(1, 17) = 126.2$, $p < 0.001$ between vowel and speaker suggests that RTs for both speakers differed only significantly for /i/ utterances (*Mdiff* = 103 ms for /i/ and *Mdiff* = 0 ms for /u/ utterances). There were no other significant effects.

2.4.1.2.3 Static-sequential presentation (Experiment 3)

The analysis revealed a significant effect of condition, $F(2, 34) = 3.7$, $p < 0.05$. (*Ms* = 460 ms, 477 ms and 486 ms for the correlated, control and orthogonal condition, respectively). Duncan's multiple range post-hoc tests ($\alpha = 0.05$) attributed the effect to a significant difference between the correlated and the orthogonal condition. There was a highly significant two-way interaction, $F(1, 17) = 57.5$, $p < 0.001$, between speaker and vowel. Reaction times were shorter for Speaker A when pronouncing /i/ compared to /u/ (483 ms vs. 461 ms), while for Speaker B the opposite was the case (463 ms vs. 491 ms, for /i/ and /u/, respectively). This effect was further qualified by a three-way interaction between condition, speaker and vowel, $F(2, 34)$, $p < 0.05$. There were no other significant effects.

2.4.1.2.4 Catch trials

Finally, planned t-tests were performed for each experiment comparing the catch-trials in the correlated condition with the trials in the orthogonal condition. For static stimuli, RTs were significantly longer in the catch-trials than in orthogonal trials, *Mdiff* = 36 ms, $t(17) = -3.7$, $p < 0.01$. For static-sequential stimuli, the difference between catch-trials and orthogonal trials approached significance, *Mdiff* = 28 ms, $t(17) = -1.9$, $p = 0.07$. No such difference was seen for dynamic stimuli, *Mdiff* = -5.8 ms, $t(17) = -0.3$, $p > 0.78$.

2.4.2 Movement onset corrected RTs

Because it is possible that vowel-specific anticipatory mouth movements that occur before visible mouth opening can contribute to speechreading, the possible contribution of such movements in Experiment 2 was also evaluated. For each of the 48 dynamic clips used in Experiment 2, two raters independently determined the first video frame that contained *any* kind of articulatory facial movement prior to mouth opening. Inter-rater reliability was good, $r = 0.81$. For each stimulus, estimates of both raters were averaged and converted into milliseconds. An ANOVA was then performed using utterance types (Speaker A /u/, Speaker A /i/, Speaker B /u/, Speaker B /i/) and clips as random variables. There was a highly significant effect of utterance type, $F(3, 44) = 6.9$, $p < 0.001$. Post-hoc testing using Duncan's multiple range test ($\alpha = 0.05$) indicated that anticipatory movements for Speaker A occurred earlier on average as compared to Speaker B (M = 116 ms vs. 61 ms prior to the first visible mouth opening). No significant differences were observed between vowels. Mean values were in order: 100 ms, 132 ms, 62 ms, and 60 ms for Speaker A /u/, Speaker A /i/, Speaker B /u/, Speaker B /i/, respectively.

As a result of these systematic speaker differences, an additional ANOVA was performed on the RTs in Experiment 2 when these were measured relative to the first articulatory movement in the stimulus, rather than relative to first mouth opening. The onset corrected RTs in Experiment 2 were also compared to RTs in Experiments 1 and 3 in an additional ANOVA across experiments. The main difference to the previous analyses was that, because the period of anticipatory mouth movement was now added to the RTs in the dynamic presentation mode, RTs were significantly slower overall than those for static–sequential stimuli, though still faster than those for static stimuli, $F(2, 51) = 25.0$, $p < 0.001$; Duncan's multiple range test at an α-level of 0.05; mean values were 675 ms, 579 ms, and 475 ms for static, dynamic, and static-sequential stimuli, respectively).

Moreover, differences between Speaker A and B, which had averaged to a highly significant difference of 51 ms when measured from mouth opening (see above) were eliminated in dynamic presentation mode, *Mdiff* = -3 ms, $F(1, 17) < 1$. This suggests that the previously reported speaker differences in the dynamic condition were in fact due to speaker differences in the timing of anticipatory mouth movements. In all other respects, this ANOVA on RTs measured to movement onset reproduced exactly the same pattern of results as the ANOVA on RTs measured from the onset of mouth opening.

2.5 Discussion

In the present experiments, it was investigated whether speeded classifications of facial speech were influenced by task-irrelevant speaker variations. A significant orthogonal interference was only found for dynamic stimuli. For static-sequential stimuli there was a significant difference between the correlated and the orthogonal condition, but not between the control and the orthogonal condition. For static pictures, the overall condition effect only approached significance and this trend seemed to be due to differences between the correlated and the orthogonal condition. As mentioned before, the interpretation of a redundancy gain in the correlated condition in the selective attention paradigm bears some constraints (see also Eimas et al., 1978; Green et al., 1991). It is possible that subjects - against task instructions - use the easier dimension for their decisions when it is correlated with a more difficult one. There is some evidence that vowel classifications are more difficult than identity judgments in speeded two-choice classification tasks (Schweinberger et al., 1998) and matching tasks (Campbell et al., 1996b), although it can be speculated that this depends heavily on characteristics represented by the particular speakers. The finding of reliably longer RTs for the catch-trials relative to the orthogonal trials for static faces and a similar trend for static-sequentially presented faces indicates that subjects may have used task-irrelevant speaker information in the correlated condition. However, this was not the case in Experiment 2 (dynamic presentation).

The results of Experiment 1 and 3 seem to contradict previous reports of influences of identity on the processing of static facial speech (Schweinberger et al., 1998). However, for the static faces in Experiment 1 the numerical difference of 13 ms between the control and the orthogonal condition was practically identical to the 14 ms difference reported by Schweinberger et al. (1998) in an experiment which did not yield a significant difference for a sample of twelve participants. In a subsequent experiment, the authors increased the number of participants to 27 and found a significant difference of

30 ms between the control and the orthogonal condition. Therefore, it cannot be completely ruled out that the absence of orthogonal interference in Experiment 1 was due to a lack of statistical power, although the number of eighteen participants clearly exceeded Schweinberger et al.'s (1998) initial sample of twelve subjects. It might also be argued that the greater number and variability of the facial speech stimuli in Experiment 1 compared to the study by Schweinberger et al. (1998) possibly increased overall RT variability, making it more difficult to find a significant condition effect. On the other hand, the significant difference of 33 ms between the control and the orthogonal condition for dynamic faces in Experiment 2, where stimulus variations where similar to Experiments 1 and 3 make it unlikely that statistical power was the only crucial factor for the absence of orthogonal interference for statically and static-sequentially presented faces. However, the nil results in Experiments 1 and 3 should still be interpreted with caution, as the interaction between the condition effect and experiment did not reach significance. This might point to the direction of overall similar condition effects in the three experiments, but the non-significant interaction might also be due to a lack of statistical power.

Another reason for the contradiction between the present results and the previously reported orthogonal difference for static stimuli (Schweinberger et al., 1998) might lie in the significantly increased stimulus set used here. In the cited study, only four stimuli per control block (one speaker, displaying two exemplars of two vowels), and eight stimuli per orthogonal block (two speakers, displaying two exemplars of two vowels each) were presented. This means that stimulus set size was well below the average memory span for verbal material in the control condition and just above that margin in the orthogonal condition (Miller, 1956). Less is known about the average memory span for faces, but it is possible that participants were able to keep all stimuli active in working memory in the control, but not in the orthogonal condition of Schweinberger et al's (1998) study. With such a small stimulus set, it cannot be ruled out that, especially in the control condition, participant's classifications were based on memorized pictorial cues, which might have been completely unrelated to facial speech cues. In the present experiment, the control condition consisted of 24 different faces instead of only four stimuli, a number that possibly exceeded the average memory span for faces. This made sure that both in the control and the orthogonal condition, participants based their decisions on the relevant speech cues and not on memorized irrelevant pictorial features. However, an explanation of Schweinberger et al's (1998) results only in terms of memory effects and picture based response strategies does not explain why they did not find an orthogonal interference of vowel variations on identity classifications. The reason might lie in a confound between the task-relevant and the task-irrelevant dimension caused by the way the Garner paradigm was applied by the authors. It has been shown that participants effectively use external features such as hairstyle in identity tasks (Ellis et al., 1979, see also section 1.2.2). With such a strategy, it might be relatively easy to ignore differences in the mouth area associated with different vowels. If, throughout the experiment, faces of two persons, including external facial features are presented, no condition effect should be expected in the identity task, because both in the control and the orthogonal condition, there is an identical number of discriminative external features (e.g. hairstyles), irrespective of additional variety in the mouth region. Importantly, the aspect of the stimulus that is used to perform the identity task remains completely unaffected by introducing a second varying dimension. The situation is very different in the facial speech task, however. There are considerable differences in the way individuals pronounce vowels, which was also clearly visible in the present experiment. Therefore, doubling the number of speakers in the orthogonal condition of the speech task also increased the number of physical “vowel exemplars” from four to eight. A reaction time increase can be expected here, because in addition to the task-irrelevant speaker variation, there was also an increase of the task-relevant vowel dimension. In principle, this was also the case in the present experiments, where 24 different stimuli had been presented in the control, and 48 in the orthogonal condition. The crucial difference is that in the present experiments there was considerably more relevant stimulus variety and less repetition in the baseline of the control condition compared to the study by Schweinberger et al. (1998), where stimulus set size in the control condition was very small. It has been demonstrated that a linear increase in stimulus variety does not necessarily lead to a linear increase in reaction times, especially for larger stimulus sets (Mullenix et al., 1990), which might explain the absence of a reliable orthogonal interference in Experiments 1 and 3. The results of Experiment 1 and 3 are therefore in line with research that suggests largely independent functional processes for static facial speech and facial identity.

However, there was a clear effect of irrelevant speaker variations on classifications of vowels for dynamic facial speech. The effect was significant both when RT measurements were taken from absolute movement onset and from the first visible mouth opening. The observed effects are in line with a recent study, reporting that the percentage of correctly speechread keywords was increased when subjects were presented with single-speaker as compared to multiple-speaker lists (Yakel et al., 2000) and extend the results to processing speed. It might therefore be that an integration of identity and facial speech information requires dynamic information.

One concern with previous studies using static pictures is that an influence of irrelevant speaker variations on speechreading might reflect differences in relative speed with which identity and facial speech might be processed. Specifically, such an effect might be observed if the irrelevant information (identity) was perceived faster than the task–relevant information (for more discussion see Garner, 1983) and both dimensions are processed in a serial manner. However, in the present study orthogonal interference caused by irrelevant speaker variation was only seen for dynamic faces, when identity information was available prior to the onset of facial speech. In this situation it can be argued that identity processing was already completed when the speechreading decision had to be made. It is therefore unlikely that the present effect of orthogonal interference can be explained by a faster processing of identity compared to facial speech.

The large number of different stimuli in the present experiments also discounts any interpretation in terms of picture based response strategies. Differences of picture set size between control and orthogonal blocks represent an unlikely explanation for the orthogonal interference in Experiment 2. If such differences mainly accounted for the RT increase even for the large stimulus sets used here, the effect should also be present in Experiments 1 and 3, which had, apart from the lack of dynamic information exactly the same design as Experiment 2. These findings might suggest that dynamic facial speech is not processed completely independently of speaker characteristics. Theoretically, orthogonal interference in the Garner paradigm can also be explained by a "normalization process", as suggested by the Bruce and Young model (1986). It puts forward the idea that all speech irrelevant information is stripped off from the face stimulus. Such a process is supposed to be time consuming and might explain processing costs in multiple speaker lists. However, there is no obvious reason why such a normalization process should be more demanding for dynamic stimuli, as non speech related differences between speakers are similar both for non-rigidly moving and static faces. Nevertheless, at the moment it cannot be differentiated between *processing costs* in the orthogonal condition due to such a normalization process or *processing benefits* in the control condition due to the usage of speaker specific representations in short term memory. Evidence for one of these possible interpretations can be gained by looking at the effects of familiarity. An overall speechreading advantage for familiar faces would argue for processing benefits due to a usage of stored identity-specific speaker characteristics. Such a finding would clearly argue against the postulated normalization process. This hypothesis will be tested in Experiments 4 and 5.

The large overall reaction time delay for the static presentation mode relative to both the dynamic and the static-sequential presentation mode, independent of experimental condition, might indicate that irrespective of task requirements, facial identity has to be taken into account automatically, resulting in substantially increased RTs for static faces. In static presentation mode, identity and speech information were given at the same time, while dynamic and static-sequential stimuli provided subjects with identity information 1000 ms prior to facial speech onset. It can be speculated that an automatic processing of identity required additional resources in the static experiment, resulting in significantly longer RTs. This result is in line with Campbell and De Haan (1998) who found a significant priming effect for identity judgments when subjects took part in a prior speech-reading task, which did not require taking the identity of the speaker into account. These and other findings point into the direction of an automatic, task-independent processing of facial identity. Nevertheless, it should be noted that alternative interpretations of the slow RTs for static stimuli cannot be completely ruled out. For example, there might be time demands for the general adaptation to the onset of a face stimulus. Also, even though in static trials it was tried to present the fixation cross close to where the relevant mouth area would subsequently appear, it cannot be completely ruled out that the onset of a static face might have caused a redirection of spatial attention or an eye movement to the relevant stimulus features, and that such processes accounted for the additional time demands observed for these stimuli.

The comparison between the dynamic and the static-sequential presentation mode underlines the role of time-dependent information even for relatively simple vowel utterances. On average, the apex of the articulation was presented 320 ms after the first visible mouth opening and 408 ms after the first movement prior to mouth opening. However, the RT differences between dynamic and static-sequential faces were considerably smaller (e.g. only between 17 ms and 104 ms, respectively). If responses were governed mainly by the apices of the articulations, longer RTs for moving faces would be expected. Apparently, the additional information inherent in mouth movements was able to compensate for this delay. The dynamic clips included additional information such as critical transitions, which were not available in static or static sequential clips. This result is in line with other work that emphasizes the role of dynamic information in speechreading (for a review see Rosenblum & Saldaña, 1998).

A related observation in this experiment was that speaker differences in RTs were significantly influenced by presentation mode, with by far the most prominent speaker differences seen for dynamic stimuli. However, subsequent analyses revealed that the speaker effect disappeared when RT measurements for the dynamic clips were taken relative to movement onset, rather than to the first mouth opening. This corresponded well with different onsets of anticipatory mouth movements in the two speakers.

Movement timing during articulations is a complex process that is sensitive to subtle differences in speaking both intra- and inter-individually (Munhall, 2001). Again, the present result underlines both the role of speaker-specific mouth movements in the early phase of an articulation (in this case, before the mouth actually opened) and the capacity of the speechreading system to use such very subtle dynamic cues.

Overall, the interaction between vowel and speaker was least pronounced in the control condition. Apparently, speaker-specific differences in vowel perception are attenuated when there is no speaker variation between trials.

In summary, the present experiments demonstrate that task-irrelevant speaker variations can influence the processing speed of facial speech. However, reliable effects of speaker variations were only observed for moving faces, contradicting previous research with static images of vowel utterances. The results emphasize the importance of time dependent information for speechreading, and even differences of articulatory movements prior to mouth opening were shown to systematically affect performance. The results do not necessarily argue for an early *integration* of facial speech and identity processing. It can be argued that some speaker characteristics are held active in working memory, enabling participants to *tune* to a particular speaker when presented in a single speaker list (see also Yakel et al., 2000, for a similar conclusion). It will be tested in Experiment 4 and 5, whether speaker characteristics of familiar faces are stored in long-term memory and can be used in order to facilitate speechreading performance.

3 Experiments 4 and 5: Influences of familiarity on speechreading speed for moving faces.

3.1 Purpose of Experiments 4 and 5

Experiment 2 demonstrated longer reaction times in a speechreading task when talker identity varied in comparison to a condition in which talker identity was held constant. Two explanations might account for the results. First, they might reflect a time-consuming normalization process, which gets rid of all task-irrelevant identity information as postulated by the Bruce and Young model (1986). Alternatively, the difference between the control and the orthogonal condition might actually be the result of a processing benefit in the single talker list. In the control condition, it might be possible to keep identity-specific characteristics in working memory and to use these to enhance speechreading. This can be tested by presenting faces of familiar talkers. Visual speech cues differ considerably between talkers (Kricos, 1996). If the speech processing system was able to store and use idiosyncratic speaker information, there should be processing benefits for highly familiar speakers. At the moment, the relationship between familiarity and speechreading efficacy is unclear and there are conflicting reports. There are reports suggesting that familiarity can influence speechreading (Schweinberger et al., 1998; Walker et al., 1995), while Campbell et al. (1996a) did not find influences of familiarity on speechreading for static faces. In Experiments 1 to 3 speaker differences were most pronounced for dynamic faces and there were significant inter–individual differences even in the timing of anticipatory mouth movements. If familiarity improves speechreading due to fast access to stored idiosyncratic information, dynamically presented speech should yield stronger effects than static speech, because dynamic speech is what we usually encounter during conversations. A finding of more efficient speechreading for familiar dynamic faces would underline the notion that speechreading is to some extent contingent on identity processing.

3.2 Rationale of Experiments 4 and 5

On the basis of the formerly described idiosyncratic speaker effects, the influence of personal familiarity on speechreading speed was tested in a second set of experiments. If facial speech perception is functionally independent from the processing of facial identity, there should be no advantage for speechreading from highly familiar faces. Overall, subjects who were unfamiliar with both speakers showed no speaker differences when RTs were measured from the first visible mouth movement. Processing advantages for familiar faces would further specify the influence of identity on facial speech processing. This was tested in Experiments 4 and 5. It was decided to use only the dynamic clips because it was assumed that these represent the most ecologically valid stimuli.

Obviously, any difference between familiar and unfamiliar speakers should decrease with an increasing number of encounters with an initially unfamiliar speaker. This is an important consideration especially in the context of the present experiments, which involved the presentation of a large number of trials. It was therefore expected that potential effects of personal familiarity on speechreading would show up most clearly in the early part of the experiment. For the analysis of effects of personal familiarity, the results were therefore broken down into three consecutive trial blocks.

As in Experiments 1 to 3, the selective attention paradigm design was used, in order to keep the experiments as comparable as possible. Reaction times were measured both from first mouth openings as well as from first visible mouth movements. In a design that investigates familiarity effects, it is crucial that overall speaker effects are not confounded with familiarity effects. It was therefore made sure that one group of participants only was familiar with Speaker A and a second group of participants only to Speaker B.

3.3 Method

3.3.1 Participants

In Experiment 4 data of twelve participants (seven women and five men) aged 23-33 years (*M* = 26.6, *SD* = 2.9 years) who were all personally familiar only to Speaker A were collected.

In Experiment 5 twelve different participants (ten women and two men) between 23 and 34 years (*M* = 28.1, *SD* = 3.3 years) took part. All of them were personally familiar only to Speaker B.

Apart from one participant in Experiment 4 and two participants in Experiments 5, all subjects were German native speakers. These three participants were native speakers of Spanish. Because /i/ and /u/ vowels do not differ significantly between German and Spanish, this was considered not to be critical. In addition, all three Spanish native speakers had lived in Germany between two and six years and were all fluent in German.

Subjects in Experiments 4 and 5 rated their familiarity to both speakers on a 7-level rating scale, with "very well known" post-hoc coded as "6" and "not known" coded as "0". (Mean ratings for Experiment 4: Speaker A: *M* = 4.4, *SD* = 1.3; Speaker B: *M* = 0, *SD* = 0; mean ratings for Experiment 5: Speaker A: *M* = 0, *SD*= 0; Speaker B: *M* = 4.4, *SD* = 1.5). Most subjects in each group were close friends to one of the speakers, with nearly daily personal contact over a period of several years. All participants reported never to have seen the other speaker before. All subjects of Experiment 4 and 5 received a fee of 15 deutsche marks (DM). In Experiment 4 one additional subject was replaced due to excessive error rates (11% as compared to an average of 2.4%).

3.3.2 Procedure

In Experiment 4 and 5 the same stimuli as in Experiment 2 (dynamic video clips) were used. Task, instruction and general stimulus presentation were identical to Experiment 2. Subjects were instructed to classify /u/ and /i/ utterances by button presses on the keyboard while disregarding the identity of the speakers.

3.4 Results

Error rates were generally low, with average percentages of errors of commission (including the catch-trials) of *M* = 2.4% in Experiment 4 and *M* = 1.1% in Experiment 5. Low error rates were observed for all experimental conditions in both experiments: (Experiment 4, *Ms* = 1.7%, 2.6% and 2.1% for the correlated, control and orthogonal conditions; Experiment 5, *Ms* = 0.9%, 1.5% and 1.1% respectively). The error rates in the catch-trials of the correlated condition were *M* = 3.3% in Experiment 4 and *M* = 1.3% in Experiment 5. Error rates were not analysed further. Outliers were very rare (*M* = 0.2% in Experiment 4 and *M* = 0.3% in Experiment 5).

Reaction times were evaluated in two different types of analysis. In the first analysis, overall effects of condition (correlated, control, orthogonal) were tested in an analogous way to Experiments 1 to 3. In the second analysis, effects of personal familiarity were evaluated separately for different levels of experimental familiarization in the three successive trial blocks. Experiment was always defined as between-subjects variable.

3.4.1 Effects of condition.

The effect of experimental condition was not significant, $F(2, 44) < 1$ (*Ms* = 456 ms, 463 ms and 469 ms for the correlated, control and orthogonal condition, respectively). As in Experiment 2, utterances of Speaker A were classified faster than utterances of Speaker B (*Mdiff* = 63 ms), $F(1, 22) = 277.0$, $p < 0.001$. This overall speaker effect was not significantly modulated by experiment, as suggested by the non-significant interaction, $F(1,22) = 2.4$, $p = 0.14$. Overall, /u/ utterances were classified faster than /i/ utterances (*Mdiff* = 27 ms), $F(1, 22) = 21.6$, $p < 0.001$. There was a significant two-way interaction between condition and speaker, $F(2, 44) = 4.3$, $p < 0.02$, indicating slightly smaller differences between the two speakers in the correlated condition. A significant interaction between speaker and vowel was observed, $F(1, 22) = 114.4$, $p < 0.001$, indicating longer RTs for /i/ utterances of Speaker B compared to Speaker A (*Mdiff* = 96 ms) and smaller RT differences for /u/ utterances (*Mdiff* = 9 ms). There were no other significant effects.

3.4.2 Analysis of trial blocks.

In order to test whether participants who were familiar with one of the speakers were able to use idiosyncratic mouth movements more efficiently, results were broken down into three consecutive trial blocks, irrespective of the counterbalanced experimental condition. In the ANOVA for the first block of 96 trials, the interaction between experiment and speaker approached significance, $F(1, 22) = 3.0$, $p = 0.09$. In the

ANOVAs of the second and the third trial blocks, no interactions between experiment and speaker were observed, $F < 1$. In all three analyses, highly significant main effects of speaker, vowel, and a significant interaction between speaker and vowel were present. However, these effects were the same as described in previous analyses and are therefore not listed here.

The interaction between experiment and speaker in the first trial block reflected a tendency of faster speechreading for familiar speakers. In Experiment 4 (participants familiar with Speaker A), the advantage for Speaker A vowels was 65 ms. In Experiment 5 (participants familiar with Speaker B), the advantage for Speaker A was reduced to 45 ms. (In the corresponding data of Experiment 2, with participants who were unfamiliar with both speakers, the advantage for Speaker A was 59 ms).

3.4.3 Analyses of movement onset corrected RTs

Similar to Experiment 2, ANOVAs were also performed on the movement onset corrected RTs using the same factors as described above.

3.4.3.1 Effects of condition

The effect of experimental condition was not significant, $F(2, 44) < 1$. There was a significant effect of vowel, $F(1, 22) = 50.5$, $p < 0.001$, indicating shorter RTs for /u/ utterances (*Mdiff* = 42 ms) and a significant two-way interaction between condition and speaker $F(2, 44) = 4.51$, $p < 0.05$. The highly significant interaction between vowel and speaker, $F(1, 22) = 44.2$, $p < 0.001$ suggested smaller differences between /u/ and /i/ utterances for Speaker A (*Mdiff* = 15 ms) compared to Speaker B (*Mdiff* = 69 ms). There were no other significant effects.

3.4.3.2 Analysis of trial blocks

In addition to the already described vowel main effect and the interaction between speaker and vowel, the analysis of the first block of onset corrected trials showed a trend for the two-way interaction between experiment and speaker, $F(1,22) = 3.6$, $p = 0.07$. According to this, participants who were familiar with Speaker A tended to speechread slightly faster from this speaker (*Mdiff* = 11 ms), and participants who were familiar with Speaker B responded slightly faster to utterances of Speaker B (*Mdiff* = -10 ms). This trend was not visible in the later blocks ($F(1, 11) < 1$ in blocks 2 and 3, respectively).

Subsequent analyses per experiment did not show significantly shorter RTs for the familiar Speaker A in Experiment 4, $F(1, 11) = 1.2$, $p = 0.29$, (*Mdiff* = 11 ms), but there was a trend for shorter RTs for the familiar Speaker B in Experiment 5, $F(1, 11) = 4.6$, $p = 0.05$ (*Mdiff* = 10 ms). Familiarity effects are illustrated in Figure 8, which also shows RT differences between both speakers for the first block of onset corrected trials in Experiment 2, where participants where unfamiliar with both speakers.

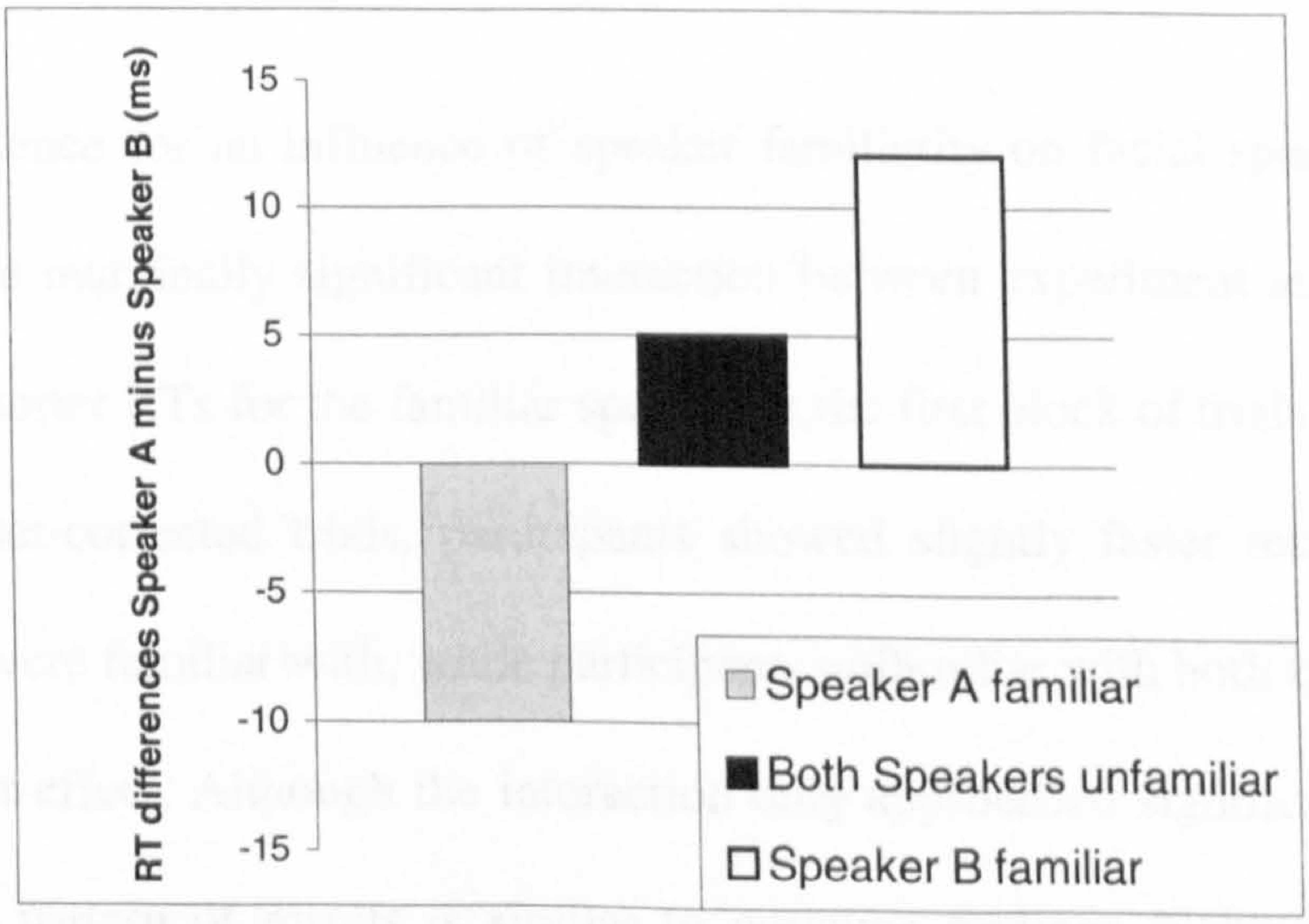

Figure 8: Reaction time differences between Speaker A and Speaker B for the first 96 movement onset corrected trials in Experiment 4 (participants only familiar with Speaker A), Experiment 2 (both speakers unfamiliar) and Experiment 5 (participants only familiar with Speaker B).

3.5 Discussion

In Experiments 4 and 5, which included speakers that were personally familiar with the participants, no significant influence of irrelevant speaker variations on speechreading speed was observed. This is in contrast to the finding of orthogonal interference caused by variations of unfamiliar faces in Experiment 2, which raises the possibility that effects of task-irrelevant speaker variations are modulated by familiarity. Possibly it is less difficult to encode facial speech for varying pairs of one familiar and one unfamiliar speaker, relative to varying pairs of two unfamiliar speakers. The absence of a condition effect is at variance with reports based on experiments with static faces (Schweinberger et al., 1998). The authors reported orthogonal interference in the selective attention paradigm also for subjects who were personally familiar with one of two speakers. However, it has already been pointed out that these results might have been biased by other factors than dependencies between facial speech and facial identity, such as picture based strategies and differences of task-relevant variability between experimental conditions.

Some evidence for an influence of speaker familiarity on facial speech processing comes from the marginally significant interaction between experiment and speaker and the trend for shorter RTs for the familiar speaker in the first block of trials in Experiment 5. For the onset-corrected trials, participants showed slightly faster responses for the speakers they were familiar with, while participants unfamiliar with both speakers did not show a speaker effect. Although the interaction only approached significance, it may be noted that the pattern of results is similar to previous findings (Schweinberger et al., 1998, but see also Campbell et al., 1996a). The overall pattern of the results makes it likely that the interaction was only marginally significant due to the relatively small number of twelve participants per experiment causing a lack of statistical power. However, because only persons who had a lot of face-to-face communication experience with one of the talkers were included, the number of potential participants was very limited. It is noteworthy that the trend for a familiarity effect was confined to the first block of trials and disappeared in subsequent blocks in which the unfamiliar speaker may have become familiar as a result of a large number of encounters. Even though the familiarity effect was only marginally significant and should therefore be interpreted with caution, these observations may add some evidence that familiarity might facilitate speechreading even in simple tasks that require the classification of single vowels only.

The finding of an influence of familiarity on facial speech processing is in line with a report of a modulation of the McGurk effect depending on speaker familiarity (Walker et al., 1995). In contrast to the present experiments and to reports by Schweinberger et al. (1998), other researchers did not find more efficient speechreading for familiar speakers (Campbell et al., 1996a; Campbell et al., 1998). A possible explanation for this might lie in the different definition of familiarity in the studies. In the present experiments and in the study by Schweinberger et al. (1998), participants were all close acquaintances and friends of the speakers and all of them used to have a lot of face-to-face communication over a period of years. The speakers in the study by Campbell et al. (1996a) were recruited from university staff, probably known only from sight to the participating students. Possibly, more direct communication experience with a particular face is necessary in order to find an impact of personal familiarity on speechreading.

Furthermore, because articulatory movements differ to a considerable extent between speakers (Montgomery et al., 1983), speechreading might be optimised by a flexible system that makes use of dynamic idiosyncratic speaker properties. In this study, the presence of such effects was consistently shown especially for anticipatory movements in dynamic stimuli, whereas idiosyncratic effects were much smaller for static and static-sequential presentation. The stimuli in the present experiment included dynamic information, which was not the case in the studies by Campbell et al. (1996a; 1998). Stored representations of idiosyncratic speaker characteristics might include critical time-dependent features (but see also Schweinberger et al., 1998). Experiments 4 and 5 also suggest that speaker specific representations can be established quickly. After about one hundred trials, speech from the formerly unfamiliar speaker was decoded with equal speed as utterances displayed by a familiar face. It is therefore unlikely to find familiarity effects if responses are averaged across a large number of trials, which include responses made when participants were already familiarized to all talkers.

The trend for an influence of familiarity on speechreading speed and the findings of Experiments 1 to 3, which demonstrated a reliable orthogonal interference for unfamiliar faces only for dynamic stimuli might suggest an early integration of dynamic speech and identity information. This argues against a strong version of the independence hypothesis and does not support the notion of a normalization process stripping off identity specific stimulus characteristics. Additional support for this interpretation comes from recent evidence that dynamic facial speech information can also be used to make judgments about speaker identity in the absence of structural cues (Rosenblum et al., 2002).

To my knowledge, the present experiments were the first to investigate influences of familiarity on speechreading speed for precisely timed dynamic stimuli. It will be interesting to see whether stronger effects of face familiarity will be found for more complex stimuli such as words or whole sentences in the future.

To summarize, it was observed that speechreading speed tended to be faster for highly familiar faces. The results provide further evidence that the processing of dynamic facial speech might make use of stored speaker specific characteristics and argue against earlier postulated normalization processes (Bruce et al., 1986).

4 Experiments 6 and 7: Asymmetric interactions between identity and expression? Controlling for an asymmetric increase of task-relevant information in the selective attention paradigm.

4.1 Purpose of Experiments 6 and 7

There are conflicting reports with respect to the relationship between face recognition and the processing of emotional expression (see also section 1.2.6). Recently, work has been published that suggested an "asymmetric interaction" between both processes (Schweinberger et al., 1998; Schweinberger et al., 1999). In these studies, the Garner paradigm of selective attention was applied and orthogonal interference was only found when expression, but not identity, was the relevant dimension. The authors concluded that identity can be accessed independently of facial expression, while the processing of expression might be partly contingent on facial identity information. However, the interpretation of their results bears some problems. In particular, the conclusions are based on observations using very small stimulus sets. The potential problems arising from that have already been outlined above (see also section 2.5). In addition to a potential influence of memory effects and picture based response strategies, the results might have been caused by differences in the increase of task-relevant information from the control to the orthogonal condition. Originally, the Garner paradigm has been used to explore the relationship between relatively basic perceptual processes, such as form and colour. With these dimensions, it is possible to add variations of the respective task-irrelevant stimulus dimension in the orthogonal condition, without affecting variability of the relevant dimension. To illustrate, in the control condition of a colour classification task participants are presented with either squares or circles and decide whether the stimuli are either blue or green. In the orthogonal condition, form is varied orthogonally to colour. This additional task-irrelevant form variation leaves the colour dimension unaffected, because both processes depend on different and physically independent features. Irrespectively of the fact that squares *and* circles are presented in the orthogonal condition, the wavelengths signalling blue and green remain exactly the same. Most importantly, this is different for such complex stimuli as faces. Both expression and identity information are mediated by at least partially overlapping physical features (see also sections 1.2.2 and 1.2.3). This means that additional variation of a task-irrelevant facial dimension (e.g. identity) might also increase the variation of the task-relevant dimension (e.g. expression). If this is not taken into account, stimulus dependent interactions might be confused with interactions between processes. Most critically, under certain conditions a potential increase of relevant information might differ between an identity and an expression task. This might occur if pictures of only two individuals displaying two exemplars of two expressions each are presented as in the studies by Schweinberger et al. (1998; 1999). Assume participants base their identity classifications simply on external features, such as hairstyle, which would be an efficient strategy under the described conditions. In the identity task, there would be no increase of task-relevant information from the control to the orthogonal condition, because doubling the amount of expression categories in the orthogonal condition leaves the number of physically different external features used for the identity task unaffected. However, this would not be the case in the expression task. Because humans show considerable variation in the way they express emotions and because especially for the analysis of expression the whole facial configuration has to be taken into account, the assumption that the relevant expression dimension is held completely constant across the control and the orthogonal condition is at least highly questionable. Doubling the number of stimuli in the orthogonal condition by presenting an additional face identity also increases task-relevant expression information, because the number of physically different expressive displays is also doubled from four to eight.

There is another serious concern with respect to very small face stimulus sets in the selective attention paradigm. If only four pictures per block are presented, it might not be certain that participants really perform a face identity task. Pictures of the same person might also be very similar in a non-identity related, superficial aspect, such as overall contrast or brightness (because e.g. pictures of one person were taken on the same day under the same lighting conditions). Then, classifications of "identity" might be made based on such cues, without the need to actually encode the identity of a face. Theoretically, such a pictorial decision might be made even without attending to the face at all (if e.g. pictures have highly distinctive flaws). In the expression task however, it is far less likely that particular expressions correlate strongly with pictorial cues such as lighting. Both expressions are displayed by both posers and overall, identity correlated pictorial cues are more likely to vary within one particular expression. Therefore, participants do have to attend to the relevant expressive features in the mouth and eye region and cannot easily make decisions on the basis of superficial pictorial cues. Following this reasoning, a stronger influence of the "irrelevant" identity dimension can be expected in the expression task because adding pictures of an additional face also increases variety of the relevant expression dimension. Under such conditions, rather than revealing the architecture of identity and expression processes, an asymmetric interaction might reflect asymmetric increases of task difficulty from the control to the orthogonal condition. It was the purpose of Experiments 6 and 7 to test whether the pattern of an asymmetric interaction between facial identity and facial expression still holds when these potential confounds are controlled for and when it is made sure that participants really use the relevant facial information in both tasks.

4.2 Rationale of Experiments 6 and 7

If overall task difficulty increases from the control to the orthogonal condition in the selective attention paradigm, a RT increase is likely to be observed. Such an effect does not necessarily have to be interpreted in terms of "orthogonal interference", because it might not reflect processing costs caused by the interference of an additional dimension. Processing costs might reflect an increase of task difficulty caused by greater variability with respect to the relevant dimension and might be independent from the relationship between two hypothesized processes. If the increase in task difficulty from the control to the orthogonal condition differs between two tasks, one might expect an asymmetric interaction of "orthogonal interference". Applied to the processing of facial identity and emotional expression, this means that reliable conclusions about the functional relationship between both processes can only be made if it is made sure that there is no difference between the control and the orthogonal condition with respect to task-relevant information. This was achieved by presenting exactly the same number of different stimuli in all blocks of all conditions in the Garner paradigm, making sure that in both tasks, conditions only differed with respect to task-irrelevant information. By presenting a large number of stimuli and introducing a considerable overall variability with respect to non-identity related pictorial information, it was also ensured that participants could not base decisions on superficial face-unrelated cues.

In Experiment 6, participants performed speeded classifications of emotional expressions and decided, whether a face was either happy or neutral. In Experiment 7, subjects decided as quickly as possible, whether a face was either familiar or unfamiliar. The decision, whether a face is familiar or not is supposed to trigger activation along the identity processing pathway (Bruce et al., 1986), including FRUs and PINs for each successfully recognized face. Therefore, experimental familiarity decisions provide a possibility to study face identity processing by presenting larger stimulus sets than the ones used in previous studies which applied the selective attention paradigm (Schweinberger et. al., 1998; Schweinberger et al., 1999). The application of a face familiarity task means that "identity variations" were defined in a slightly different way from the cited studies where identity decisions were related to two individual faces. Here, identity variations were defined with respect to variations between a "familiar" and an "unfamiliar" category, not with respect to individual within-category variations.

Importantly, in order to prevent picture based response strategies, a large overall stimulus set of 160 pictures was used. The pictures consisted of 40 familiar and 40 unfamiliar faces, displaying a neutral and a happy expression each. The large number of individual faces and the fact that pictures were not repeated within experimental conditions made sure that familiarity decisions could not be made solely by attending to repeatedly presented identical external features. Most importantly, the number of different stimuli per block and condition was identical in both tasks. As a consequence, there was no increase of relevant information from the control to the orthogonal condition in both tasks.

If the processing of emotional expression is contingent on facial identity, as suggested by Schweinberger et al. (1999) a similar interaction as previously described (Schweinberger et al., 1998; Schweinberger et al., 1999) should be found. If the reported asymmetric interaction was produced by asymmetric differences in task difficulty, it should disappear in the present design.

4.3 Method

4.3.1 Stimuli and Apparatus

Identical stimulus sets were used in both experiments. Pictures of forty male celebrities displaying a happy and a neutral expression each were selected from a newspaper's archive (*Südkurier Konstanz*, Germany). The photographs were scanned using an AGFA Snapscan1212^{TM}. For each celebrity, an unfamiliar face of similar general appearance and age was matched. This resulted in a stimulus set of 160 pictures (40 happy familiar faces; 40 neutral familiar faces; 40 happy unfamiliar faces and 40 neutral unfamiliar faces). Photographs of unfamiliar faces were taken from various sources with the intent to obtain a similar degree of superficial stimulus variability such as lighting, contrast and overall picture quality as in the familiar set. All pictures were digitally edited using Adobe Photoshop^{TM}. All background was removed and an attempt was made to equalize contrast and brightness. Examples of the stimuli can be seen in Figure 9.

The stimuli were presented on black background in the centre of a 19'' monitor that was connected to an IBM compatible personal computer. The presentation software was ERTS^{TM} (Experimental Runtime System, Berisoft Corporation). Picture resolution was 17.7 pixels/cm at a screen resolution of 800 by 600 pixels The size of the stimuli was 6 cm x 7.6 cm at a resolution of 28.3 pixels/cm. Viewing distance was 60 cm, resulting in a horizontal visual angle of 5.7 degrees and a vertical visual angle of 7.2 degrees.

4.3.2 Procedure

In all trials a white fixation cross on a black background was shown for 500 ms, followed by a face stimulus visible for 1500 milliseconds or until a key was pressed. After a key-press, the face disappeared and there was a blank screen for 1000 ms. Visual feedback in form of the words "too slow!" or "too fast!" (in German) was only given 1500 ms after stimulus onset for missing and slow (reaction times > 1200 ms) or extremely fast answers (RT < 150 ms).

After reading the instructions on the monitor, participants were shown stimulus examples that were not used in experimental trials. Both experiments consisted of three experimental conditions, labelled "correlated", "control" and "orthogonal". Conditions consisted of two blocks containing 80 different stimuli each. There were no stimulus repetitions within conditions. Overall, the stimulus sets per condition were identical. The order of conditions and the order of blocks within conditions was completely counterbalanced across participants. Responses were made with both hands by key presses on a standard computer keyboard using the left and right "Ctrl" keys. The assignment of response hand to response alternative was completely counterbalanced across participants.

In Experiment 6, participants classified in a speeded forced two-choice task, whether the presented faces either displayed a happy or a neutral expression. In the control condition, the task-irrelevant familiarity dimension was held constant: in one block, there were only familiar faces, displaying happy and neutral expressions while in the other block only unfamiliar happy and neutral faces were shown. In the correlated condition, there was a strong co-variation between familiarity and expression: in one block, 90% of the familiar faces were showing a happy and 90% of the unfamiliar faces were displaying a neutral expression. In the other block, 90% of the unfamiliar faces were happy and 90% of the familiar faces were neutral. In order to discourage response strategies using familiarity despite task instructions as response criteria, 10% of the trials in the correlated condition served as “catch-trials”, for which the co-variation between familiarity and expression was reversed. In both blocks of the orthogonal condition, half of the 40 familiar and unfamiliar faces were presented showing a happy expression and the other half was displaying a neutral expression. In the second block of the orthogonal condition, the faces who had been presented happy now showed a neutral expression and vice versa.

Participants in Experiment 7 classified, whether the presented face was either familiar or unfamiliar. In the control condition, the expressions of the faces were held constant: in one block, there were only happy faces, while in the other block only neutral faces were shown. In one block of the correlated condition, 90% of the familiar faces were happy and 90% of the unfamiliar faces were displaying neutral expressions. In the other block, 90% of the unfamiliar faces were happy and 90% of the familiar faces showed a neutral expression. Ten percent of the trials in the correlated condition served as “catch-trials”. In both blocks of the orthogonal condition, half of the 40 familiar and unfamiliar faces were presented with a happy expression and the other half was displaying a neutral expression. In the second block of the orthogonal condition, the faces that had been presented happy now showed a neutral expression and vice versa.

After completing the tasks, participants rated on a 7-level rating scale, with "seen very often" post-hoc coded as "6" and "never seen before" coded as "0", how familiar they were to the celebrities′ faces before taking part in the experiment. In order to make sure that they were not accidentally familiar with any of the supposedly unknown faces, participants were shown the happy exemplar of each unfamiliar face and asked, whether they had seen the respective person before. No subject had seen any of the unfamiliar faces before.

First an ANOVA across experiments with repeated measurement factors on the variables condition (correlated vs. control vs. orthogonal), relevant dimension (happy vs. neutral in the expression task and familiar vs. unfamiliar in the identity task) and irrelevant dimension (happy vs. neutral in the identity task and familiar vs. unfamiliar in the expression task) were performed, in order to test for a possible "asymmetric interaction" between experiment and condition. Then, separate analyses for each experiment were performed with the repeated measurement factors condition (correlated, control, orthogonal), familiarity (familiar vs. unfamiliar) and expression (happy vs. neutral). Catch-trials were not entered into this initial analysis. When performing ANOVAs, Epsilon corrections for heterogeneity of covariances, where appropriate, were performed with the Huynh-Feldt method (Huynh-Feldt, 1976) throughout, and α–levels for post-hoc ANOVAs were Bonferroni corrected. Only answers between 150 and 1500 ms were analysed.

4.3.3 Participants

Twelve participants (eight woman and four men) aged 19–26 years ($M = 23.9$ years, $SD = 5.9$) contributed data in Experiment 6. The average familiarity rating for the celebrities was $M = 4.1$, ($SD = 0.8$). Data for one additional participant was replaced due to low familiarity ratings ($M < 2.5$).

Twelve different participants (eight women and four men) aged 20 – 34 years (M = 25.8, SD = 5.0) contributed data in Experiment 7. The average familiarity rating for the celebrities was M = 4.5 (SD = 0.9). Data of two additional participants had been replaced due to excessive error rates (M > 30% in at least one experimental condition compared to an average across participants of M = 8%).

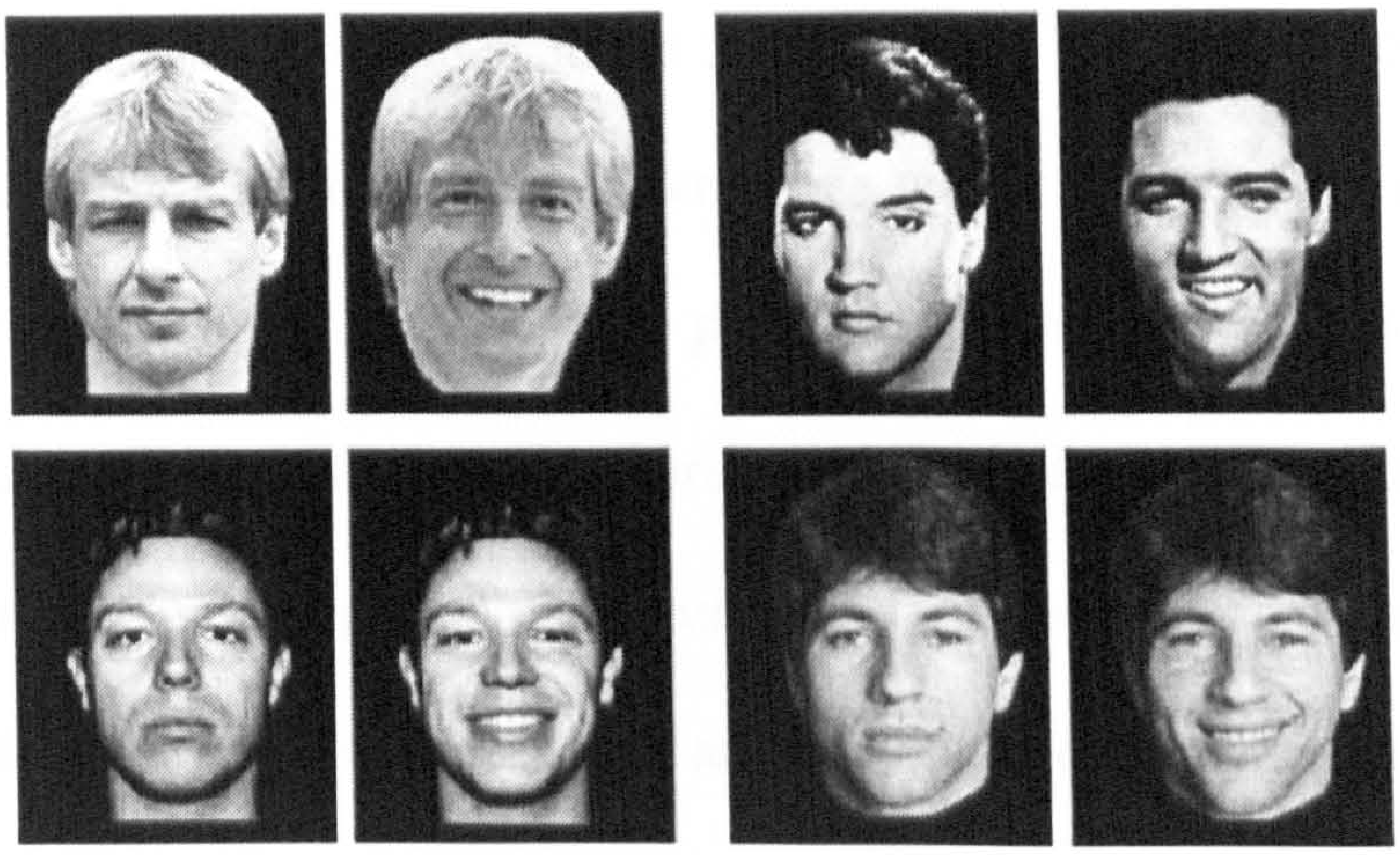

Figure 9: Examples of stimuli used in Experiments 6 and 7. Top row: familiar faces (from left to right: Jürgen Klinsmann and Elvis Presley) neutral and happy, respectively. Bottom row: matched unfamiliar faces.

4.4 Results

Missing and invalid answers were extremely rare (M < 0.1% in both experiments) and were not analysed further.

4.4.1 Reaction Times

The analysis across experiments did not reveal significant differences between the two tasks, $F(1,22) < 1$. Importantly, there was no effect of condition, $F(2, 44) < 1$ and no interaction between experiment and condition, $F(2, 44) < 1$ (see also Figure 10).

The separate analysis for the expression task (Experiment 6) yielded no significant effect of condition, $F(1, 11) < 1$. Overall, expressions displayed by unfamiliar faces were recognized slightly faster than expressions displayed by familiar faces, $F(1, 11) = 5.6$, $p < 0.05$ ($M = 598$ ms vs. $M = 609$ ms). This effect was further qualified by a significant two-way interaction between expression and familiarity, $F(1, 11) = 33.2$, $p < 0.001$, which suggested that for unfamiliar faces, both expressions were classified with similar speed, while for familiar faces there seemed to be an advantage for happy expressions. The interaction was further qualified by a significant three-way interaction between condition, expression and familiarity, $F(2, 22) = 9.0$, $p < 0.01$. These interactions were further explored by separate ANOVAs for each experimental condition (Bonferroni corrected α-level = 0.017), including the repeated measurement factors expression (happy vs. neutral) and familiarity (familiar vs. unfamiliar). The analyses yielded highly significant two-way interactions between expression and familiarity for the correlated $F(1,11) = 45.7$, $p < 0.001$ and for the orthogonal condition, $F(1,11) = 17.8$, $p < 0.01$. In both cases they seemed to be due to faster classifications of happy expressions displayed by familiar faces in contrast to faster classifications of neutral expressions displayed by unfamiliar faces. For the control condition, there were no significant effects (see also Figure 11).

The analysis of RTs in Experiment 6 yielded no other significant effects.

The separate analysis for the identity task (Experiment 7) did not reveal an effect of condition, $F(2, 22) < 1$, (see also Figure 10). Overall, familiar faces were classified faster than unfamiliar faces, $F(1, 22) = 18.4$, $p < 0.01$ ($M = 611$ ms vs. $M = 643$ ms). No other main effects or interactions approached significance.

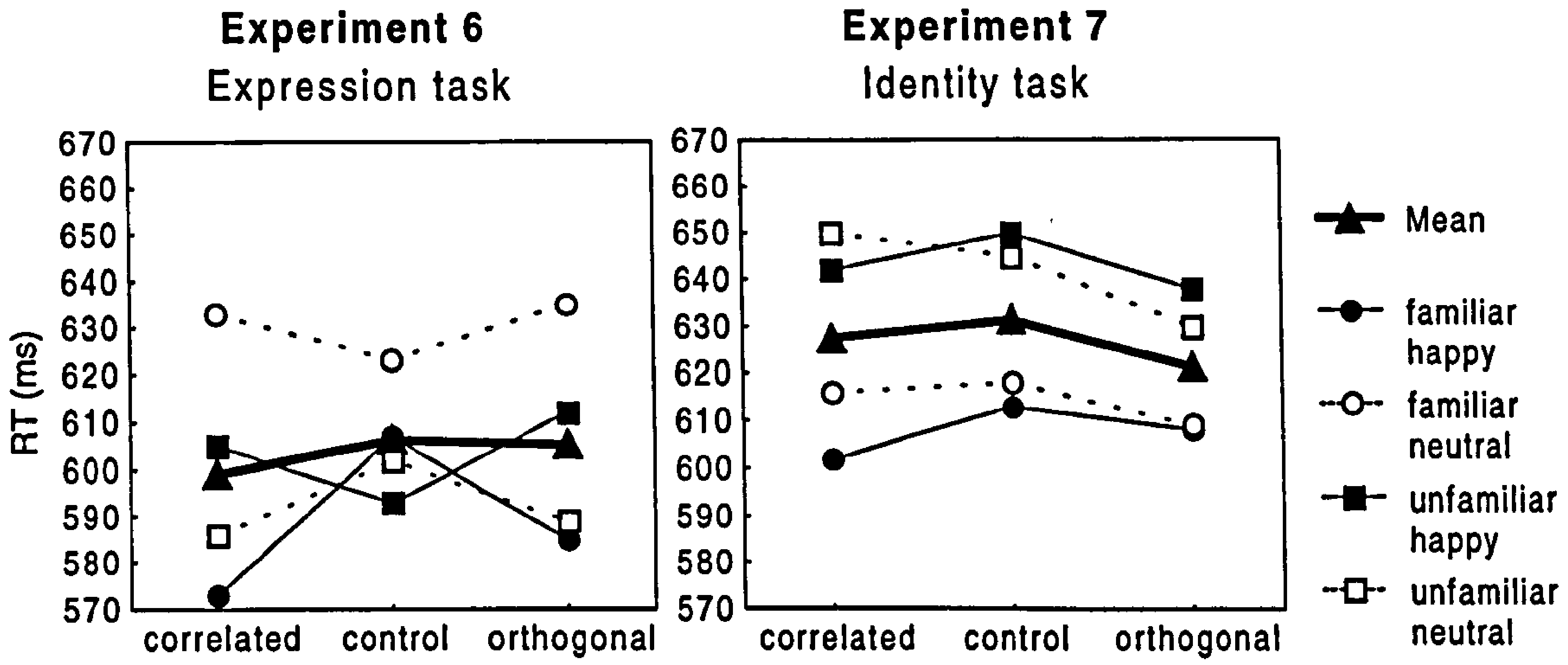

Figure 10: Mean RTs in Experiments 6 and 7. Neither the effects of condition nor the interaction between experiment and condition were significant.

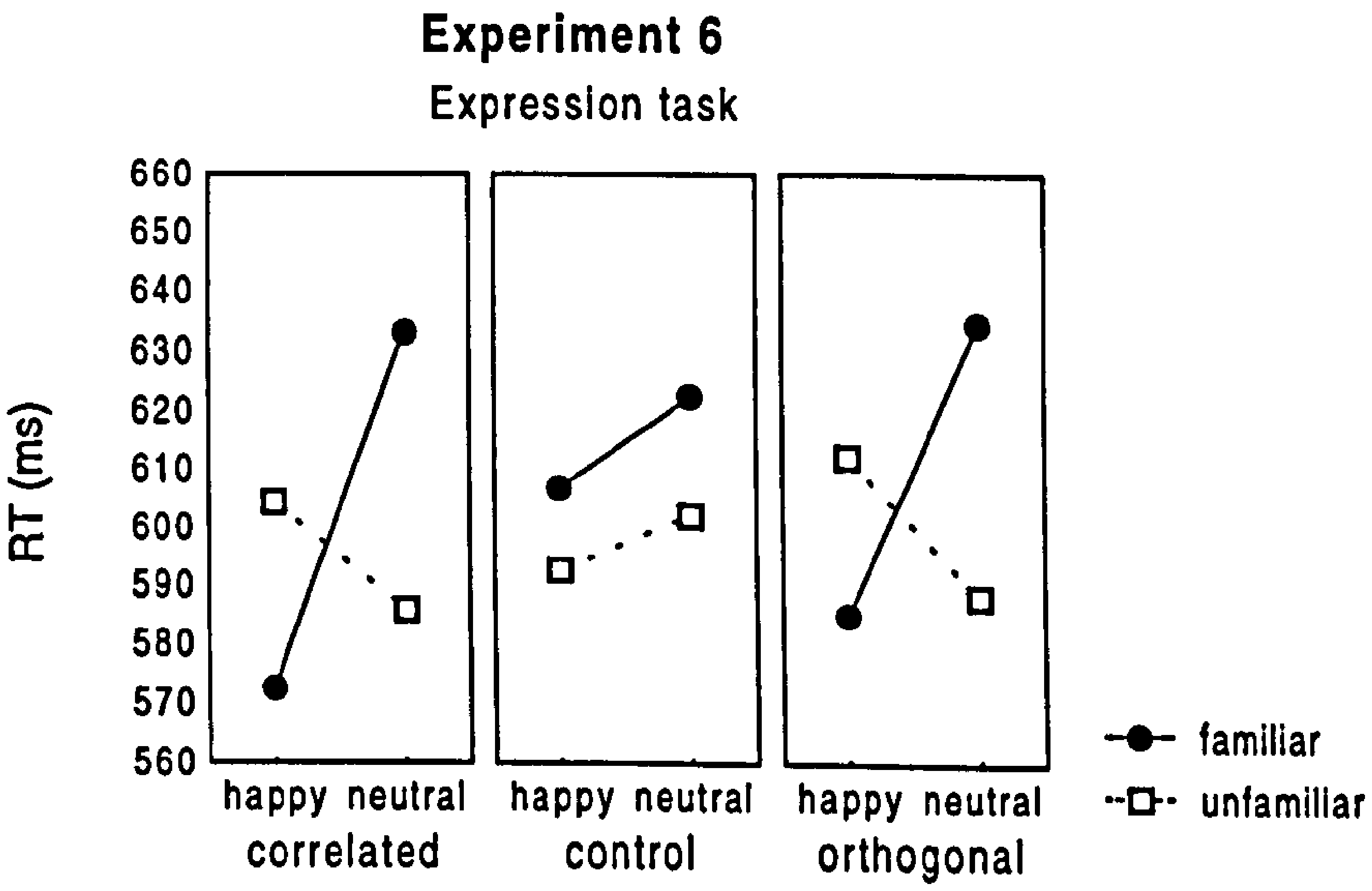

Figure 11: RTs for Experiment 6: three way interaction, $F(2, 22) = 9.0, p < 0.01$ between condition, familiarity and expression.

4.4.2 Error rates

The ANOVA across experiments did not show significant effects of experiment, $F(1, 22) < 1$, or condition, $F(2,44) = 1.6, p > 0.2$. The interaction between the two factors was not significant, $F(2, 44) < 1$.

The separate analysis for the expression task (Experiment 6), revealed no speed-accuracy trade-off. There was a two-way interaction between expression and familiarity, $F(1, 11) = 58.5$, $p < 0.001$ (see also Figure 12) and a three-way interaction between condition, expresion and familiarity, $F(2, 22) = 7.9$, $p < 0.01$ (see also Figure 13). These were further explored by separate ANOVAs per condition (α–level = 0.017), with the repeated measurement factors expression (happy vs. neutral) and familiarity (familiar vs. unfamiliar). The two-way interactions between expression and familiarity reached significance in the correlated, $F(1, 11) = 35.1$, $p < 0.001$, in the control, $F(1, 11) = 12.8$, $p < 0.01$, and in the orthogonal condition $F(1, 11) = 49.5$, $p < 0.001$.

In the identity task (Experiment 7), error rates were higher for familiar faces, $F(1, 22) = 16.7$, $p < 0.01$ ($M = 11\%$ vs. $M = 5.1\%$). This effect was further qualified by a significant two-way interaction between familiarity and expression $F(1, 11) = 5.6$, $p < 0.05$. Inspection of Figure 12 suggests this was due to slightly more accurate responses to happy familiar and neutral unfamiliar faces.

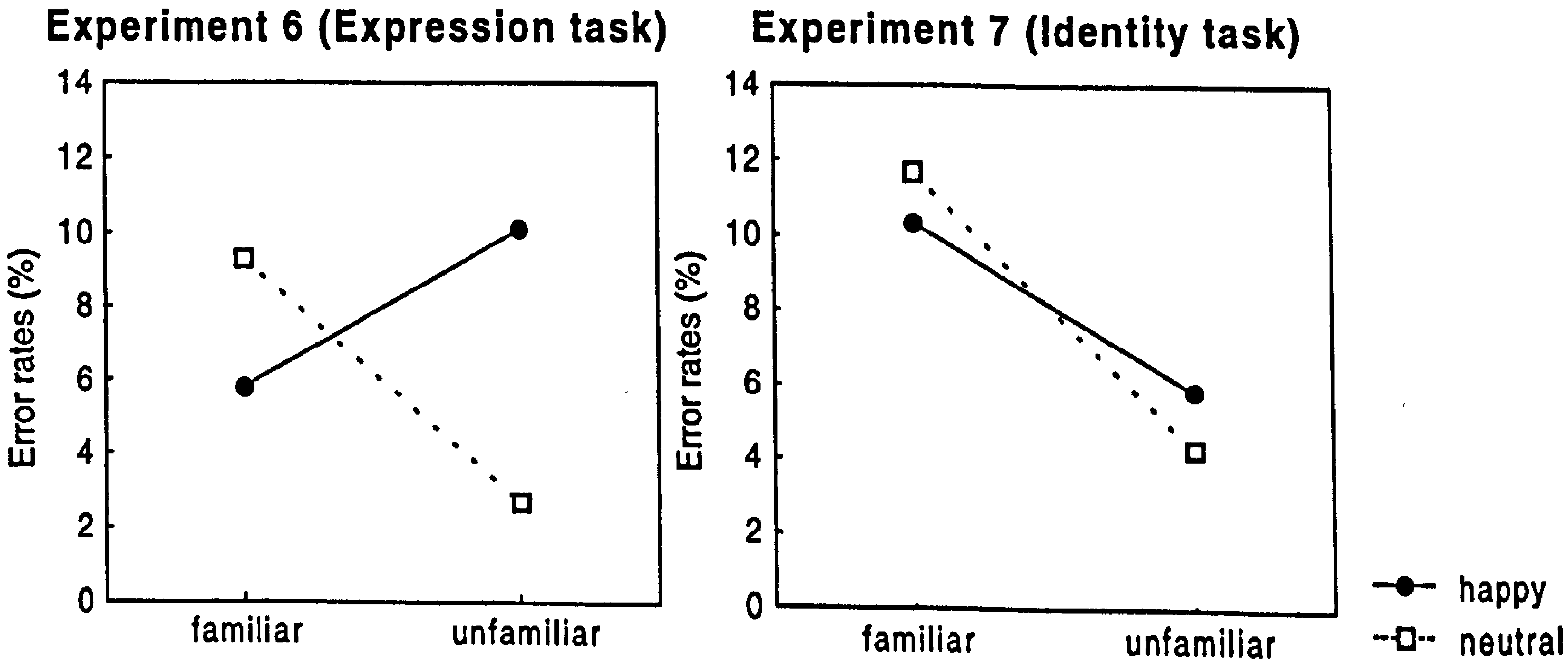

Figure 12: Error rates: two-way interactions between familiarity and expression in Experiments 6, $F(1, 11) = 58.5, p < 0.001$ (left) and Experiment 7, $F(1, 11) = 5.6, p < 0.05$.

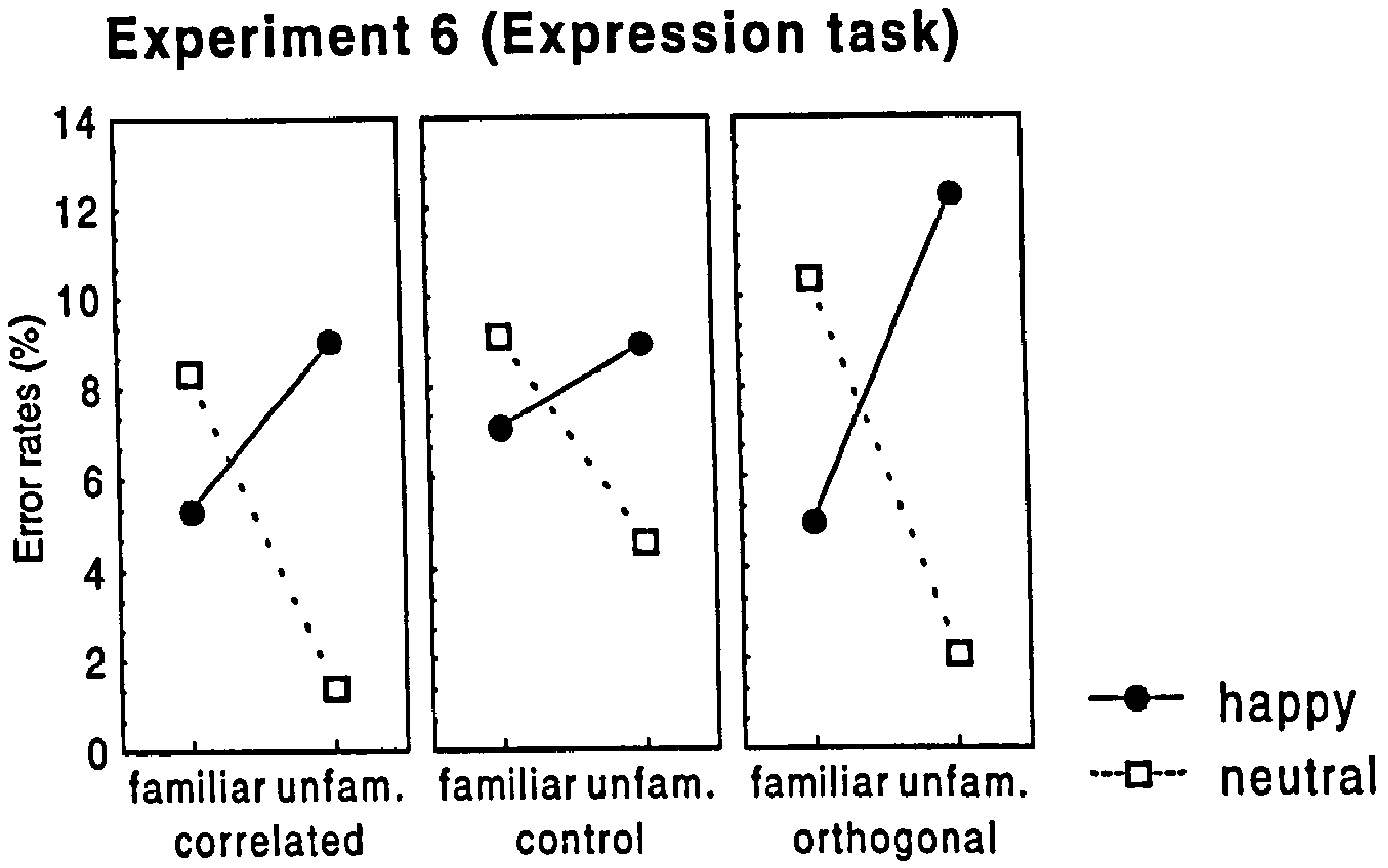

Figure 13: Error rates in Experiment 6: three-way interaction between experimental condition, familiarity and expression, $F(2,22) = 7.9, p < 0.01$.

4.5 Discussion

Recently, an asymmetric interaction between the processing of identity and emotional expression in a Garner type speeded classification task has been reported (Schweinberger et al., 1998; Schweinberger et al., 1999). The authors suggested that identity can be processed independently of task-irrelevant variations of emotional expressions, while the processing of expression might be contingent on facial identity. Using a similar selective attention paradigm, it was investigated whether the described interaction is still found when picture based response strategies or an asymmetric increase of task difficulty from the control to the orthogonal condition can be ruled out. Compared to the cited studies, the stimulus set was significantly increased and it was made sure that the control and the orthogonal condition only differed with respect to the respective task-irrelevant dimension, while the number of different stimuli presented per block was held constant across all conditions. In line with the results of Schweinberger et al. (1998;

1999), it was found that decisions on identity were not influenced by irrelevant variations of expression.

However, no asymmetric interaction between familiarity and expression decisions was found and no orthogonal interference of task-irrelevant identity variations was observed in the expression task. In contrast to the studies by Schweinberger et al. (1998; 1999), there were no stimulus repetitions within experimental conditions, ruling out stimulus based response strategies and memory effects. Here, stimulus sets for the control and the orthogonal condition were considerably larger and most importantly, of equal size. The design ruled out that there was overall more variety of relevant information in the expression task, which might have been the case in the study by Schweinberger et al. (1998), possibly making the expression task slightly more difficult than the identity task in their experiments. Differences in overall task difficulty might have an influence on orthogonal interference in the Garner paradigm. Importantly, in the present experiments both tasks were equally difficult as suggested by similar reaction times. The absence of orthogonal interference in both experiments does not support an interpretation in the sense of an asymmetric interaction between face recognition and expressions processing as suggested by Schweinberger et al. (1998; 1999). Overall, the results are in line with the notion of an independent processing of facial identity and expression (Bruce et al., 1986).

However, it has to be noted that apart from controlling the abovementioned problematic factors in the Garner paradigm, the design of Experiments 6 and 7 also differed from the cited study with respect to the way in which "identity variations" were defined. In the studies by Schweinberger et al. (1998; 1999), pictures of one individual were presented in the control condition of the expression task, while in the orthogonal condition faces of two individuals were shown. Identity was thus defined in terms of *individual face identity*. In the present study, it was face familiarity that was either held constant or varied block-wise in the expression task. Therefore, both in the control and in the orthogonal condition, a number of different individual faces was shown and identity was defined in the sense of a *super-ordinate familiarity category*. One might argue that the absence of orthogonal interference in the expression task could be due to the fact that both in the control and in the orthogonal condition a variety of individual faces was presented. Although the familiarity dimension was held constant in the control condition, there might have been interference caused by individual within category variations leading to smaller differences between the control and the orthogonal condition compared to the studies by Schweinberger et al. (1998; 1999). As in the cited experiments, the number of individual face identities was also doubled from the control to the orthogonal condition in the present expression task, so that some increase of orthogonal interference might still be expected if expression processing was contingent on identity information. However, it has been shown that orthogonal interference does not necessarily have to increase in proportion to the increase of task-irrelevant variation (Mullenix et al., 1990).

In contrast to the non-significant condition effects which argues for an independent processing of identity and facial expression, there is some evidence for an integration of both dimensions. In the expression task, there was a highly significant two-way interaction between familiarity and type of expression in blocks containing familiar and unfamiliar faces. These interactions suggested that happy expressions were classified faster and more accurately when displayed by familiar faces while the opposite seemed to be the case for neutral expressions. This finding contradicts results of expression matching tasks which have been reported to be independent of face familiarity (Bruce, 1986; Young et al., 1985). Several explanations for an influence of familiarity on classifications of expressions seem to be possible. The most trivial explanation is that the effect might reflect differences between familiar and unfamiliar faces with respect to the expressiveness of happy emotions and "non-expressiveness" of neutral faces. The celebrities might display positive emotions more convincingly for various reasons (e.g. practice). The effect then might be due to characteristics of the expression displays, instead of being primarily related to familiarity. This question can not be answered here and would have to be tested with participants who are unfamiliar with all faces. Unfortunately, this was not possible for practical reasons. However, an observation that argues against an interpretation in terms of stimulus based differences is the significant three-way interaction between condition, familiarity and expression both for reaction times and error rates in Experiment 6. Importantly, in the control condition, where the familiarity dimension was held constant, no interaction between expression and familiarity was found for the RTs, while both in the correlated and in the orthogonal condition it was highly significant. Because the interaction was not present in the baseline, one might argue that in the correlated and in the orthogonal condition, were familiar and unfamiliar faces varied within blocks, participants seemed to associate happy expressions with familiar and neutral expressions with unfamiliar faces. Might this bias have been caused by a top-down modulation of expression modules via PINs or by strategic decisions governed by the "cognitive system" (Bruce and Young, 1986)? If we assume a strict parallel processing of identity and expression as proposed by Bruce and Young (1986), ruling out interconnections between both pathways before the output of the hypothesized modules finally reaches the "cognitive system", identity processing would have to be completed faster in order to exert an influence on expression decisions. However, there was no overall RT difference between the expression and the identity classifications of expressions seem to be possible. The most trivial explanation is that the effect might reflect differences between familiar and unfamiliar faces with respect to the expressiveness of happy emotions and "non-expressiveness" of neutral faces. The celebrities might display positive emotions more convincingly for various reasons (e.g. practice). The effect then might be due to characteristics of the expression displays, instead of being primarily related to familiarity. This question can not be answered here and would have to be tested with participants who are unfamiliar with all faces. Unfortunately, this was not possible for practical reasons. However, an observation that argues against an interpretation in terms of stimulus based differences is the significant three-way interaction between condition, familiarity and expression both for reaction times and error rates in Experiment 6. Importantly, in the control condition, where the familiarity dimension was held constant, no interaction between expression and familiarity was found for the RTs, while both in the correlated and in the orthogonal condition it was highly significant. Because the interaction was not present in the baseline, one might argue that in the correlated and in the orthogonal condition, were familiar and unfamiliar faces varied within blocks, participants seemed to associate happy expressions with familiar and neutral expressions with unfamiliar faces. Might this bias have been caused by a top-down modulation of expression modules via PINs or by strategic decisions governed by the "cognitive system" (Bruce and Young, 1986)? If we assume a strict parallel processing of identity and expression as proposed by Bruce and Young (1986), ruling out interconnections between both pathways before the output of the hypothesized modules finally reaches the "cognitive system", identity processing would have to be completed faster in order to exert an influence on expression decisions. However, there was no overall RT difference between the expression and the identity task, making such an explanation unlikely. Numerically, the expression task was even completed faster than familiarity decisions. Therefore, an early interconnection of units signalling familiarity and units involved in extracting the expressive content of a face might account for the effect.

Similarly, there is evidence from Experiment 7 that familiar faces were recognized more accurately when displaying a happy expression, which might reflect that FRUs may preserve some sort of expression information. Recently, there have been reports pointing into this direction (Baudouin et al., 2000). Assuming that we are more likely to encounter faces of celebrities with smiling expressions, such a pattern might best be explained by perceptual learning.

The finding of shorter RTs for familiar faces in the identity task is in line with a number of studies (e.g. Bruce, 1986; Campbell et al., 1996a; Young et al., 1986) and is usually explained by the automatic activation of FRUs, which are only available for familiar faces. The slightly higher error rates for familiar faces probably do not reflect a speed-accuracy trade off and might be due to conservative response strategies and the fact that not all participants were familiar with all celebrities.

To summarize, the absence of orthogonal interference both in Experiments 6 an 7 argues for an independent processing of expression and facial identity in the sense of the Bruce and Young model (1986). Previous findings suggesting that face identity can be accessed independently of task-irrelevant variations of emotional expression (Schweinberger et al., 1998; Schweinberger et al., 1999) were confirmed and extended to familiar faces. In Experiment 7, where participants categorized faces as either familiar or unfamiliar, classifications were not influenced by irrelevant expression variations. The results significantly extend previous findings because memory effects and pictorial response strategies could be ruled out by presenting a large number of individual faces. Opposed to recent reports (Schweinberger et al., 1998; Schweinberger et al., 1999), there was no evidence for an influence of task-irrelevant identity variations on expression classifications. It is not completely clear whether the contrasting result was due to a more efficient control of potentially problematic factors in the Garner paradigm, such as pictorial memory effects, overall differences and asymmetric increases in task difficulty. Alternatively, a different definition of “identity variations” might account for the conflicting results. In spite of an absence of orthogonal interference in both tasks, there is some evidence that expressions are integrated to some extent with information about face familiarity, as suggested by faster and more accurate classifications of happy expressions for familiar and neutral expressions for unfamiliar faces. Similarly, slightly lower error rates for happy familiar faces in the identity task suggests that stored representations of familiar faces on the FRU level might preserve some information about frequently encountered expressions displayed by a particular face identity.

5 Experiments 8-10: Dependencies between the processing of facial identity and emotional expression? Experiments with morphed faces

5.1 Purpose of Experiments 8-10

No orthogonal interference of task-irrelevant expression or identity variations was found in Experiments 6 and 7, which applied a selective attention paradigm (Garner, 1974). This finding is at variance with previous reports of an asymmetric interaction between both dimensions (Schweinberger et al., 1998; Schweinberger et al., 1999), but is in line with the notion of parallel and independent processing of identity and expression as suggested by Bruce and Young (1986). The major aim of the following experiments was to investigate the influence of task-irrelevant stimulus manipulations on the processing of both facial identity and facial expression using a different approach. Morphing can be used to selectively manipulate stimulus salience of either facial identity or emotional expression (Beale et al., 1995; Calder et al., 1996). One prediction that can be derived from the Bruce and Young model (1986) is that manipulations of stimulus salience on the identity dimension should not influence the processing of emotional expression. Similarly, performance in an identity task should be independent of expression changes. For unfamiliar faces, this was demonstrated in a study by Schweinberger et al. (1999). Morphing from Person A to Person B did not significantly interfere with speeded classifications of emotional expression and morphing from a happy to an angry expression had no influence on the performance in an identity task.

However, studying face perception only with unfamiliar faces bears some important limitations. As there do not exist FRUs for unfamiliar faces, the independence of stored facial identity representations from emotional expressions can only be investigated using familiar stimuli. Also, evidence has emerged that face perception for familiar and unfamiliar faces differs in various aspects. Humans are very good at recognizing and matching familiar faces, and this is to a large extent independent of stimulus quality. However, for unfamiliar faces performances in matching tasks drop dramatically when lighting, view point or expression varies (for a review see Hancock et al., 2000). It has also been shown that the information we use to recognize faces differs depending on the degree of familiarity. External features such as hairstyle and changeable aspects such as a beard are especially important for the perception of unfamiliar faces, while internal and stable characteristics play a crucial role for the perception of familiar faces (Young et al., 1985).

Of particular interest for the present study was, whether morph manipulations of identity influence the processing of emotional expression and whether the recognition of familiar faces can be modulated by selectively manipulating emotional expression.

5.2 Rationale of Experiments 8-10

If facial identity and facial expression were processed in a completely independent manner, familiar faces should be recognized with similar speed and accuracy for all emotional expressions. Similarly, the performance in expression classification tasks should not depend on the degree of familiarity with a face. If, however, the expression processing system does make use of idiosyncratic identity information, which is only available for familiar faces, expressions should be classified faster and more accurately, when they are displayed by familiar faces. In order to test these predictions, the morphing technique, which allows for a selective manipulation of either facial identity or expression in realistic stimuli was used. Identity was manipulated by morphing from a familiar with an unfamiliar face within a given expression. The expression of a face was varied by morphing from a happy to an angry emotion while identity was held constant. Apart from replicating earlier findings of discontinuous classification functions with continuous stimulus changes that had been reported for morphs across facial identities (Beale et al., 1995, Schweinberger et al., 1999) or across emotions (Calder et al., 1996; Young et al., 1997; Schweinberger et al., 1999) the experiments had two major aims. Schweinberger et al. (1999) have demonstrated that morphing along a task-irrelevant expression dimension did not influence performance in an identity classification task when all face stimuli were unfamiliar. Here it was investigated whether these results extend also to familiar faces. A modulation of face recognition performance by emotional expression for familiar faces would be an indication for an at least partially integrated processing of facial identity and facial emotion and would argue against a normalization process that removes expression information as part of the face recognition pathway (Bruce et al., 1986). The study by Schweinberger et al. (1999) has also shown that morphing across two unfamiliar faces does not affect classifications of constant emotional expressions. In the present study it was tested whether classifications of expression can be modulated by morphing from a familiar with an unfamiliar face. Again, an influence of the degree of familiarity on the performance in the expression task would argue against a complete independence of the processing of identity and emotional expression.

5.3 General Method

5.3.1 Stimuli and Apparatus

Experiments 8 and 9 were conducted at the University of Konstanz, Germany and Experiment 10 was conducted at the University of Glasgow, Scotland.

Stimuli in all experiments were identical. They were based on morphs taken from digitised pictures of sixteen male faces. Eight of the presented persons were actors, sportsmen or politicians who are very familiar in Germany, but not necessarily German (for names see appendix). In order to create face-pairs, an unfamiliar counterpart matched for age and general appearance was selected for each familiar face. All faces were displaying happiness and anger, resulting in 32 original pictures. The photographs of celebrities were obtained from a newspaper archive (Südkurier Konstanz) and two raters selected pictures, which displayed unambiguous emotional expressions.

The photographs were scanned using an AGFA Snapscan1212^{TM} scanner. Pictures of unfamiliar faces were taken with a reflex camera, developed on paper and also scanned.

Posers were instructed to remember a particular situation in which they felt either anger or happiness and to express the emotion. About fifteen pictures were taken from each unfamiliar face and the same two raters again choose the version with the most convincing display of the expression. All original photographs were digitally re-edited in order to standardize size, brightness, contrast and background. The pictures were saved as greyscale bitmaps.

Based on these two sets of original photographs, two types of morph stimuli for each face pair were produced using the commercial Gryphon Morph^{TM} software (Version 2.5). *Expression morphs* were obtained by transforming happy into angry faces within a constant identity. *Identity morphs* were produced by transforming a familiar into an unfamiliar face within a given expression. In order to obtain morph stimuli of photographic quality, a large number of 570 reference points was used. These reference points were distributed in a standardized way over different facial areas such as mouth, nose, chin, eye region and outline (see also Figure 3).

Including the original pictures, one morph continuum consisted of eight steps (subsequently termed *morph levels*) with the proportions of 100:0, 86:14, 71:29, 57:43, 43:57, 29:79, 14:86 and 0:100 between the initial and the final images. The morphs were also saved as bitmaps of the same size and resolution as the original pictures.

This procedure resulted in the following four morph-continua for each of the eight face-pairs (see Figure 14 for examples): from a familiar to an unfamiliar face with a happy expression (identity morphs/happy); from a familiar to an unfamiliar face with an angry expression (identity morphs/angry); from a happy to an angry expression for a familiar face (expression morphs/familiar); from a happy to an angry expression for an unfamiliar face (expression morphs/unfamiliar). The complete stimulus set of 224 faces consisted therefore of eight face pairs with 28 stimuli each (six times four morph stimuli plus the original four pictures per face pair).

The stimuli were presented on black background in the centre of a 19'' monitor that was connected to an IBM compatible personal computer. The same computer and monitor were used for the experiments in Germany (Experiments 8 and 9) and Scotland (Experiment 10). The presentation software was ERTS™ (Experimental Runtime System, Berisoft Corporation). Picture resolution was 17.7 pixels/cm at a screen resolution of 800 by 600 pixels The size of the stimuli was 10 cm x 7.5 cm at a viewing distance of 60 cm, resulting in a vertical visual angle of 9.5 degrees and a horizontal visual angle of 7.1 degrees.

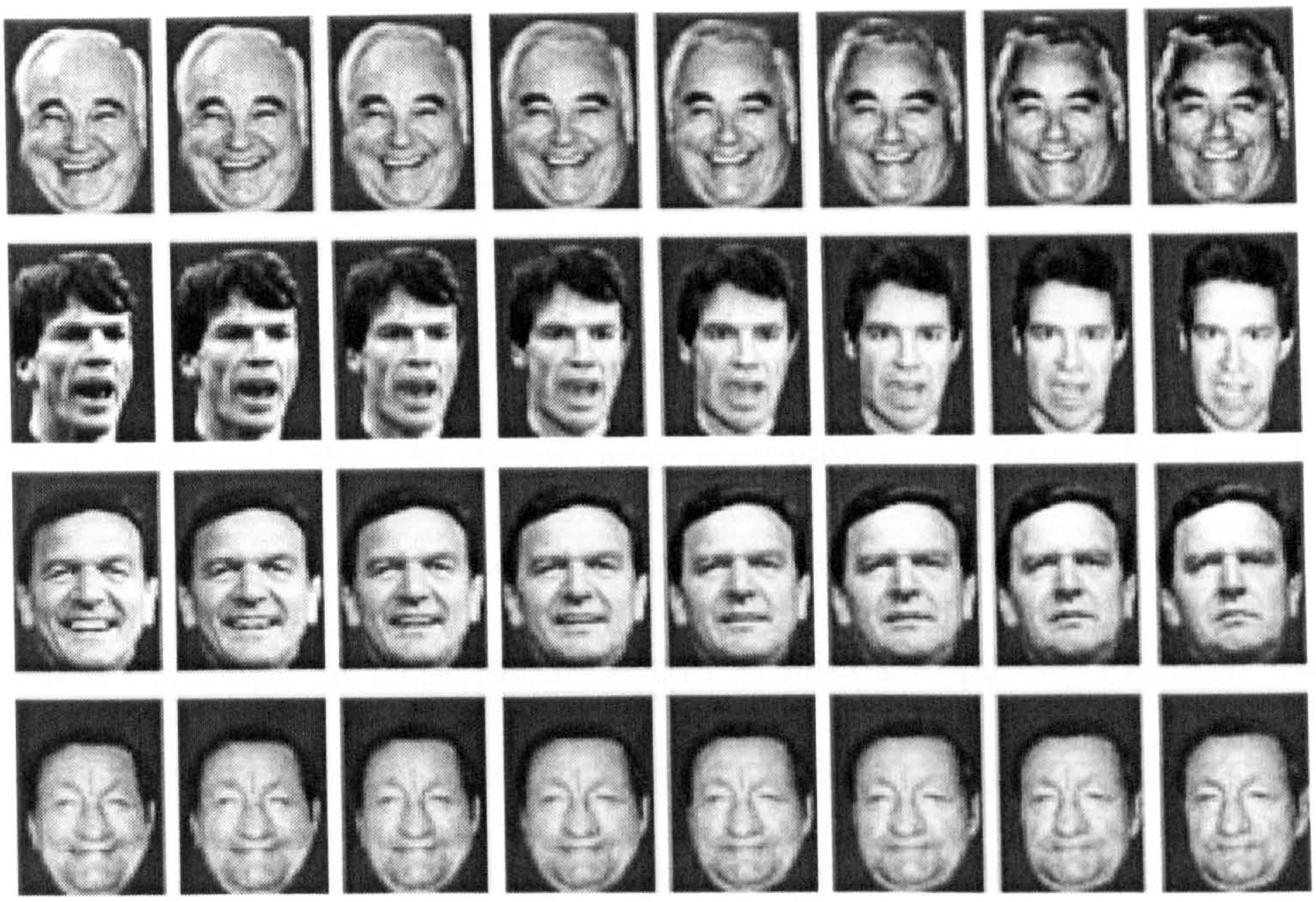

Figure 14: Examples of morph stimuli in Experiments 8-10. Top rows: identity morphs from a familiar to an unfamiliar face for happy and angry expressions. Bottom rows: expression morphs for a familiar and an unfamiliar face.

5.3.2 Procedure

Presentation of the stimuli was identical in all three experiments. In all trials a white fixation cross on a black background was shown for 500 ms, followed by a blank screen for 100 ms and a face stimulus visible for 1500 milliseconds or until a key was pressed. After a key-press, the face disappeared and there was a blank screen for 1000 ms. Feedback of a 500 Hz tone that was presented for 150 ms was only given for missing and slow answers (reaction times > 1400 ms). Both speed and accuracy were stressed. Participants responded by pressing the left and right "Ctrl" keys on a standard computer keyboard using both hands. The assignment of left and right hand responses to the particular response alternatives was counterbalanced across participants.

After reading the instructions on the monitor, participants performed twelve practice trials consisting of stimuli that were not used in the experiments and they were given the possibility to ask questions thereafter. Each experiment consisted of four blocks. Each stimulus was presented once per block so that all pictures were presented four times during the entire experiment. All experimental blocks were preceded by four additional practice trials. Practice trials were not analysed.

Within each block, the 224 stimuli were presented in random order. Each block was followed by a break. The end of the break was self-paced. The duration of the experiment was about 40 minutes. In Experiments 8 and 9, after completing the task, participants rated on a 7-point rating scale how often they had seen the eight celebrities before taking part in the study, with "seen very often" post-hoc coded as "6" and "never seen before" coded as "0". Only participants with an average score above 2.5 were included. In order to ensure that all participants were unfamiliar with all of the supposedly unknown faces, they were presented with one photograph of each unfamiliar face.

Data were averaged across face pairs, resulting in a maximum of 32 trials per condition.

When performing ANOVAs, Epsilon corrections for heterogeneity of covariances, where appropriate, were performed with the Huynh-Feldt method (Huynh-Feldt, 1976) throughout, and α–levels for post-hoc ANOVAs were Bonferroni corrected. Only answers between 150 and 1500 ms were analysed.

5.4 EXPERIMENT 8

5.4.1 Methods

5.4.1.1 Participants

Seventeen participants (eleven women and six men) aged 20-31 years (M = 22.9 years, SD = 3.2 years) contributed data in Experiment 8. The Experiment was conducted at the University of Konstanz, Germany. Participants received either a fee of 7.50 deutsche marks (DM; *n* = 13) or course credit (*n* = 4). Data from one additional participant was excluded from the analysis due to excessive error rates in at least one experimental condition where identity was not manipulated (>25%, compared to an average across participants of *M* = 2.2%). The average familiarity rating for the celebrities' faces was *M* = 4.6 (*SD* = 1). None of the participants had seen any of the unfamiliar faces before.

5.4.1.2 Procedure

Participants were informed that they would be either presented with a face out of a group of eight celebrities or an unfamiliar face. To avoid systematic false classifications, the celebrities were named before the experiment. It was pointed out that no other familiar faces than these eight would be shown. Participants decided in a speeded two-choice task whether the face was familiar or unfamiliar. Both speed and accuracy were stressed. Responses were made by pressing the left and right "Ctrl" keys on a standard computer keyboard using both hands. The assignment of left and right hand responses to familiar or unfamiliar faces was counterbalanced across participants.

5.4.2 Results

Overall, 97% of the happy and 96% of the angry original familiar faces were classified correctly. Similarly, 99% of the happy unfamiliar and 99% of the angry original unfamiliar faces were classified correctly. Missing answers were extremely rare (< 0.2%) and were not analysed further.

For the identity morphs, ANOVAs were performed on classifications and response times (RTs) with repeated measurements on the variables expression (happy vs. angry) and morph level (familiar to unfamiliar, in eight steps). Due to the ambiguous nature of the identity morphs in the identity task, no ANOVA was performed on errors of commission for these stimuli.

For the expression morphs, ANOVAs were performed on errors of commission and RTs with repeated measurements on the variables familiarity (familiar vs. unfamiliar) and morph level (happy to angry in eight steps).

Incorrect answers were not entered into the RT analyses. Answers were considered as wrong if a face on any level of the expression morph continuum was incorrectly classified as familiar or unfamiliar. On the identity continuum, wrong answers for the non-ambiguous morph levels 1, 2, 7 and 8 were excluded.

5.4.2.1 Morphs along identity (expression constant):

5.4.2.1.1 Reaction times

As expected, there was a morph level main effect $F(7, 112) = 32.4$, $p < 0.001$, with maximum RTs for intermediate morph levels. There was no effect of expression $F(1, 16) < 1$ and no interaction between expression and morph level, $F(7, 112) = 1.5$, $p > 0.18$.

5.4.2.1.2 Classifications

Expression had no overall influence on classifications of familiarity, $F(1, 16) < 1$. The morph level effect, $F(7, 112) = 772.7$, $p < 0.001$ demonstrated relatively sharp category boundaries (see also Figure 15). There was a trend for the interaction between expression and morph level, $F(7, 112) = 2.32$, $p < 0.07$. Inspection of Figure 15 suggests that on the familiar end of the morph continuum, happy faces were slightly more likely to be classified as familiar. However, separate ANOVAs on each morph level comparing happy and angry expressions were not significant (Bonferroni corrected $\alpha = 0.006$).

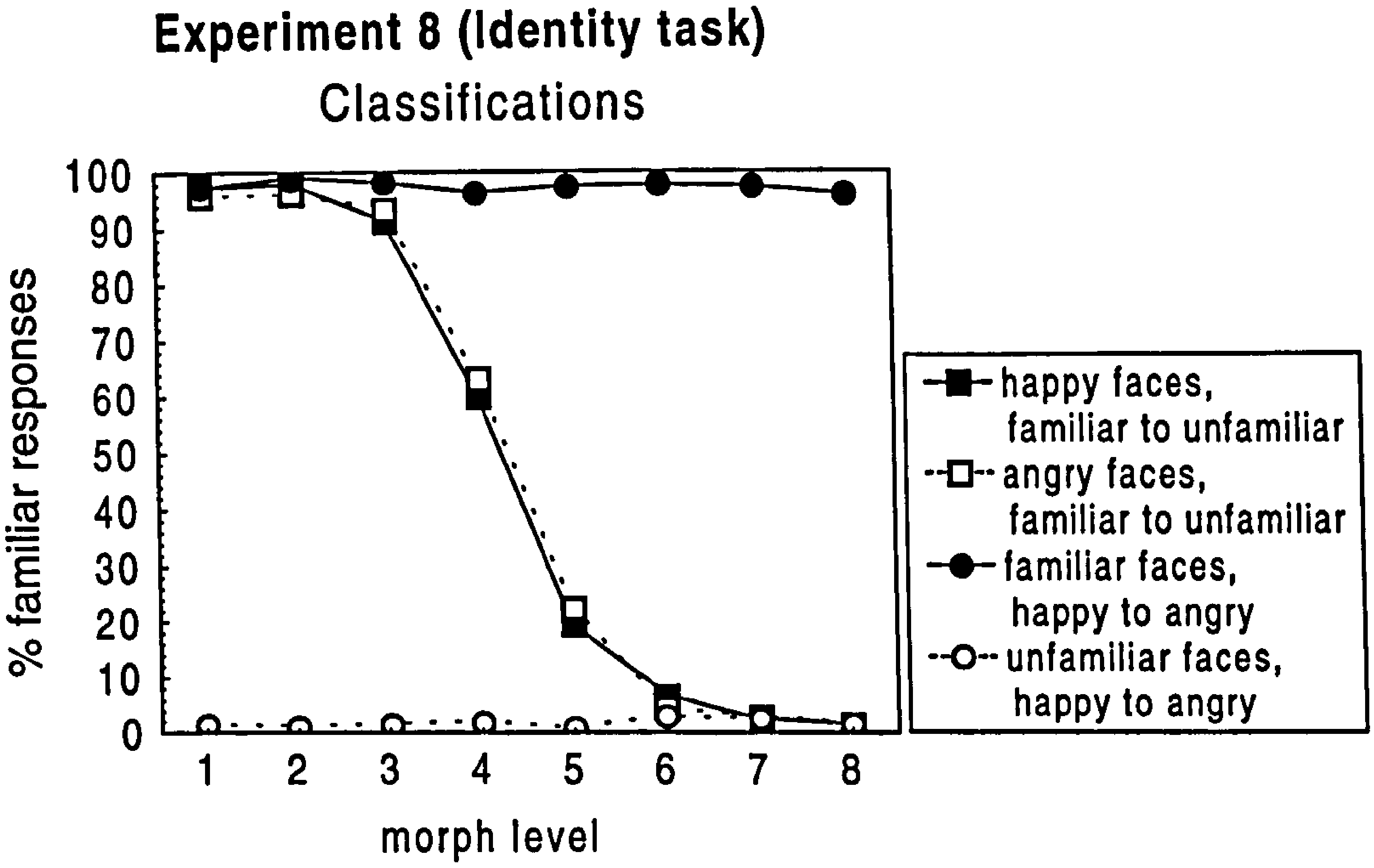

Figure 15: Results of Experiment 8. Percentage of "familiar" classifications depending on morph level and facial expression. Square symbols correspond to data for faces morphed from familiar to unfamiliar within one expression; circles correspond to faces morphed from a happy to an angry expression within a constant identity.

5.4.2.2 Morphs along expression (identity constant):

5.4.2.2.1 Reaction times:

There was an effect of morph level, $F(7, 112) = 2.2, p < 0.05$. Most importantly, this main effect was further qualified by a highly significant two-way interaction between familiarity and morph level, $F(7, 112) = 4.7, p < 0.001$ (see Figure 16), which was further explored by separate ANOVAs for familiar and unfamiliar faces. The analyses demonstrated a highly significant morph level effect for familiar, $F(7, 112) = 6.1, p <$

0.001, but no effect for unfamiliar faces, F < 1. Inspection of Figure 16 suggests that although RTs to familiar faces were shorter for the original happy faces as compared to the original angry faces, the relationship between morph level and RT was U-shaped rather than monotonic. That is, the shortest RTs were seen at the intermediate morph level 3 which corresponds to moderately happy expressions.

To evaluate this impression, orthogonal polynomial contrasts across morph levels were calculated. In addition to a linear trend, $F(1, 16) = 13.9$, $p < 0.01$, there was a quadratic trend, $F(1, 16) = 21.3$, $p < 0.001$, with no significant contribution from any higher order trends. Two planned comparisons were then performed between the RT minimum and the two RT maxima. These revealed that familiar faces with a moderately happy expression (morph level 3) were recognized more quickly than the original images of these faces with a happy expression, $F(1, 16) = 5.9$, $p < 0.05$, and were also classified more quickly than the original angry faces, $F(1, 16) = 27.9$, $p < 0.001$.

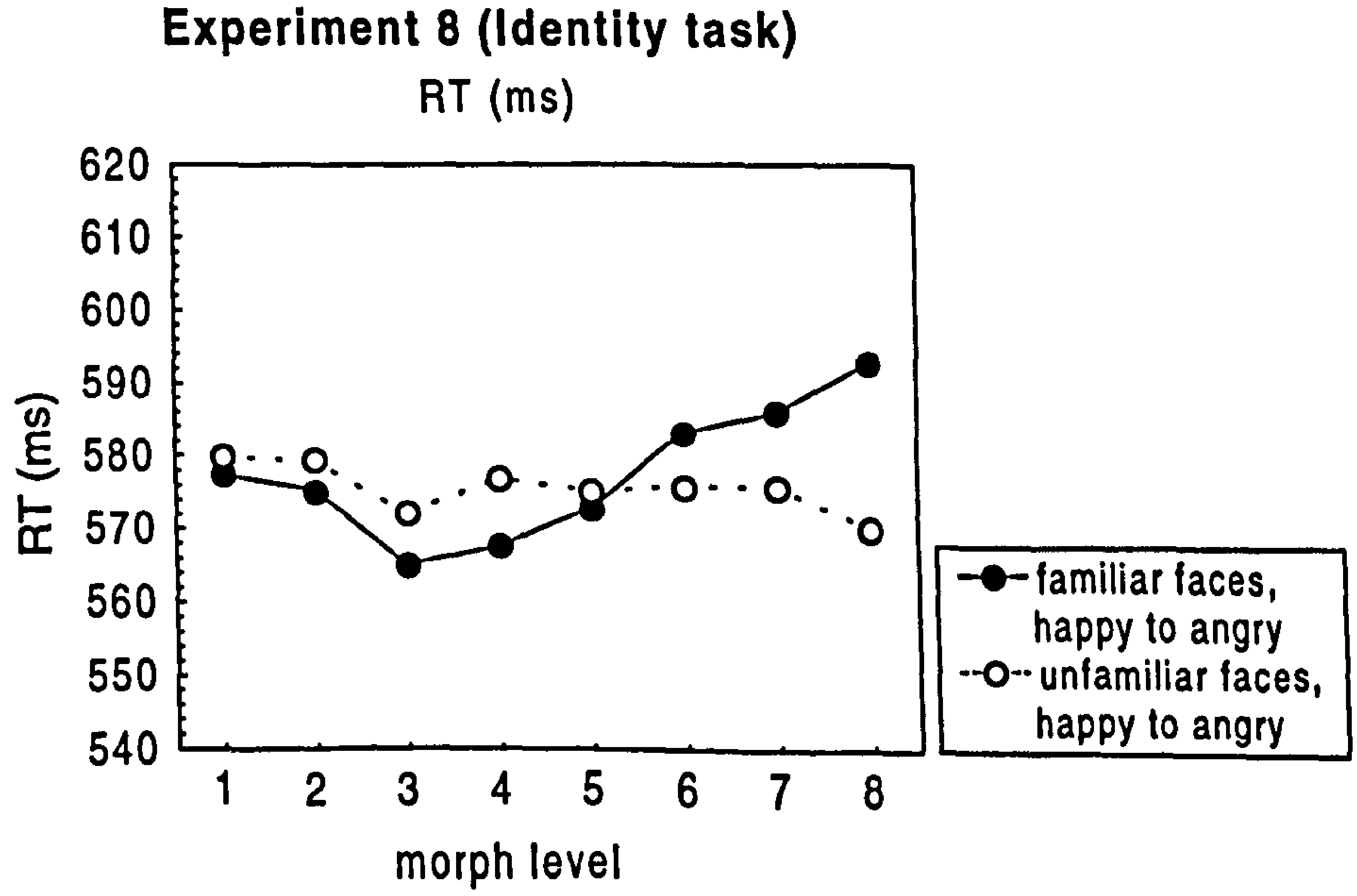

Figure 16: Classification response times for morphs along emotional expressions in Experiment 8. Note that morphing across the irrelevant expression dimension only affects RTs for familiar faces.

5.4.2.2.2 Error rates:

The analysis of errors of commission revealed a significant main effect of familiarity, $F(1, 16) = 4.86$, $p < 0.05$, reflecting slightly higher error rates for familiar faces (2.6%) in comparison to unfamiliar faces (1.8%). There was a trend for the variable morph level, $F(7, 112) = 2$, $p < 0.08$ and a trend for the interaction between familiarity and morph level, $F(7, 112) = 1.9$, $p = 0.10$.

Separate ANOVAs (at a Bonferroni corrected α-level of 0.025) for familiar and unfamiliar faces showed a trend of morph level only for familiar faces, $F(7, 112) = 2.3$, $p = 0.04$, with minimum error rates for smiling familiar faces (morph level 2, $M = 0.9\%$) and maximum error rates for angry familiar faces (morph level 8, $M = 4.2\%$).

Error rates for the unfamiliar faces did not differ significantly between emotional expressions, $F(7, 112) < 1.3$, $p > 0.25$.

5.4.3 Discussion

The present data are in line with previous findings of relatively narrow category boundaries for classifications of identity (Beale et al., 1995) and increasing perceptual difficulty for stimuli with increasing distance to the endpoints of the identity morph continuum (Schweinberger et al., 1999).

Typically, familiarity increases processing speed in identity matching or classification tasks (e.g. Bruce, 1982; Valentine et al., 1986; Young et al., 1986). Such a pattern is predicted by the Bruce and Young model (1986) because structural information for familiar faces is thought to be represented in long-term memory in form of domain specific "Face Recognition Units" (FRUs) which in turn activate non-specific "Person Identity Nodes" (PINs) that signal familiarity (Burton et al., 1990). Both FRUs and PINs are supposed to work fast and automatically, while for the rejection of unfamiliar faces directed and supposedly more time consuming visual processing is required. In this experiment, there was no overall difference between classifications of familiar and unfamiliar faces. This might be due to the fact that participants knew which faces would appear throughout the experiment, reducing uncertainty for "unfamiliar" decisions. Also, unfamiliar counterparts were very closely matched to the familiar faces, which might have encouraged conservative strategies and possibly slowed down classifications of faces as "familiar".

In line with Bruce and Young's model (1986) and Schweinberger et al.'s (1999) data, morphing along the task-irrelevant expression dimension influenced neither RTs nor error rates for unfamiliar faces. However, classification response times for familiar faces were clearly affected by morphing along emotional expressions. Participants recognized familiar faces faster when these were displaying happy compared to angry expressions. The results also suggest that especially moderately happy familiar faces were recognized fastest. A similar trend was visible for the accuracy of classifications, indicating that this pattern was not the result of a speed-accuracy trade off. This pattern might reveal information about the nature of stored facial representations rather than reflecting an effect of on the level of pictorial encoding. The influence of pictorial cues in face recognition tasks has been demonstrated in a study by Bruce (1982). When view or expressions differed between study and test phase, RTs both for familiar and unfamiliar faces were slowed down.

For three reasons, it is highly unlikely that the present results were caused by the use of pictorial cues instead of reflecting the structure of long-term stored facial representations. First, in this experiment all familiar and unfamiliar faces were shown with equal frequency. Second, RTs were shorter and error rates tended to be smaller for artificial morph stimuli which participants had never seen before taking part in the experiment, ruling out that they had any pictorial cues available for the familiar faces. Third, care was taken not to select very typical, "iconic" portraits of the celebrities as the basis of the morph pictures, (such as e.g. the famous portrait of Che Guevara). It seems therefore reasonable to speculate that the effect does not originate from the structural encoding level but from a later stage of processing.

According to Bruce and Young (1986), view and expression independent information is stored at the level of FRUs for every face we are familiar with. It is assumed that repeated encounters with a novel face lead to storage of invariant facial characteristics that are "normalized" with respect to emotional expression. This means that the face recognition system is expected to discard all information, which is irrelevant for extracting the identity of a face. The present data do not support this aspect of the model. There are at least two possible explanations for the faster RTs observed for smiling familiar faces, while classifications of unfamiliar faces were not influenced by expressions. In terms of an interactive activation model of face perception (Burton et al., 1990), top down influences from the "Semantic Information Units" (SIUs) might have a facilitating influence on PINs, if a familiar person is associated with a particular mood or expression. Activation coming from the bottom-up direction via FRUs and spreading to the PINs, where the familiarity decision is thought to be taken (Burton et al., 1990) might require less time to reach a threshold that signals familiarity if a pre-activating top-down influence is present. However, assuming a strictly parallel model, this would require a faster processing of the expression route compared to identity. This seems unlikely, as the opposite has been described for matching tasks using faces including the external facial features, (Campbell et al., 1996a; Münte et al., 1998; Potter et al., 1997; Strauss et al., 1981). In a study, where only internal features were presented and the identity task was therefore made more difficult, similar RTs for identity and expression matching were found (Bobes et al., 2000). In Experiment 9, it will be tested whether the stimuli used in Experiment 8 can be classified faster with respect to emotional expression than identity.

A further, and possibly more plausible explanation for faster recognition of smiling familiar faces might be an influence of a frequent pairing of a particular face with a certain emotional expression during face familiarization. We might be less likely to encounter celebrities with angry expressions in the media (an idea that is confirmed by the difficulty to find appropriate stimuli!). Such frequency effects might have an influence on the representations of faces at the FRU level. Structural information which is available for familiar faces might be stored in memory together with information about "typical" and “untypical” emotional expressions, resulting in better recognition of familiar faces displaying typical expressions and an inferior performance for atypical expressive displays. In this experiment, the unfamiliar faces for which no FRUs had been available prior to the experiment, were classified independently of emotional expressions. In can be speculated that during the experiment FRUs were established also for these faces as a result of frequent exposure, but because all expressions were shown with an equal frequency no effect was observed for prior unfamiliar faces. It remains to be tested, whether under experimentally controlled conditions the pairing of new face identities with a particular expression during a face learning phase leads to better recognition of faces that are presented at test displaying this expression, even if different exemplars are used.

At first sight, the results contradict a number of experimental studies that did not find an influence of emotional expressions in identity matching tasks (e.g. Campbell et al., 1996a; Young et al., 1986). However, to my knowledge, no pictures of angry celebrities have been presented in any identity matching or face recognition tasks so far, which might explain that no influence of "untypical" expressions on identity processing has been found yet.

As a general limitation that applies both to previous research and the present study, it should be noted that generalizations about a complete independence of face recognition from emotional expression based on research using a limited number of expressions may well be premature. Recent data suggest that different basic emotions are based on separate neural systems (Blair et al., 1999; Calder et al., 1996; Sprengelmeyer et al., 1997; Sprengelmeyer et al., 1998). Thus, the possibility needs to be considered that each of these systems might interact differently with face recognition areas.

After demonstrating influences of emotional expressions on classifications of familiar faces, it will now be explored whether classifications of emotional expressions can also be modulated by familiarity. Furthermore, a comparison of overall performances in the identity task with an expression classification task allows for a testing of the above mentioned top-down influences via semantic information units on familiarity processing. For such an influence to occur in a parallel and modular system, mean reaction times for classifications of emotional expressions should be faster than classifications of identity.

5.5 EXPERIMENT 9

5.5.1 Method

5.5.1.1 Participants

Eighteen different participants (twelve women and six men) aged 19-26 years (M = 21.8 years, SD = 2.7 years) took part in Experiment 9. Participants received either a fee of 7.50 deutsche marks (DM; n = 13) or course credit (n = 5). The Experiment was conducted at the University of Konstanz, Germany. The mean familiarity rating for familiar faces was $M = 4.5$ ($SD = 1$).

One subject recognized one of the unfamiliar faces and her data were replaced by an additional participant. Data from another additional participant had been replaced because of problems in sustaining concentration.

5.5.1.2 Procedure

Stimuli and presentation were identical to Experiment 8. The task required participants to make speeded two-choice classifications and decided, whether the faces displayed either a happy or an angry expression.

5.5.2 Results

Overall, original happy familiar faces were classified correctly to 96% and original happy unfamiliar faces yielded 92% correct responses. Both familiar and unfamiliar original angry faces were classified correctly to 91%. Missing answers were extremely rare ($M < 0.2\%$) and were not analysed further.

For both morph types ANOVAs with repeated-measurement factors that were identical to the ones in Experiment 8 were performed. For the expression morphs, ANOVAs were performed on the classifications of the stimuli as happy or angry and on classification response times. Due to the ambiguous nature of the expression morphs in this task, errors of commission were not analysed for this morph type.

For the identity morphs, ANOVAs were performed on errors of commission and RTs.

Only correct answers were entered into RTs analyses. Answers were considered as wrong if an expression on any level of the identity morph continuum was not classified correctly. On the expression continuum, wrong answers for the non-ambiguous morph levels 1, 2, 7 and 8 were excluded.

5.5.2.1 Morphs along identity (expression constant):

5.5.2.1.1 Reaction times:

There was a significant main effect of expression, $F(1, 17) = 27.7$, $p < 0.001$, suggesting faster RTs for happy compared to angry expressions ($M = 613$ ms vs. $M = 679$ ms, respectively). Importantly, there was a highly significant morph level main effect, $F(7, 119) = 9.2$, $p < 0.001$, which was further modulated by type of expression, as suggested by the significant two-way interaction, $F(7, 119) = 2.6$, $p < 0.05$ (see Figure 17).

Consecutively performed separate ANOVAs for happy and angry expressions yielded a highly significant effect of morphing from familiar to unfamiliar faces for happy expressions, $F(7, 119) = 8.7$, $p < 0.001$. There was also a significant main effect for angry expressions, $F(7, 119) = 4.1$, $p < 0.001$.

Overall, inspection of Figure 17 suggests a linear increase of RTs along the identity morph continuum, however, the curves for happy and angry faces slightly differ with respect to a potential quadratic trend for angry faces. Therefore, the morph level main effect and the interaction between morph level and expression were further explored by performing an analysis of polynomial contrasts, which tested for linear and quadratic trends.

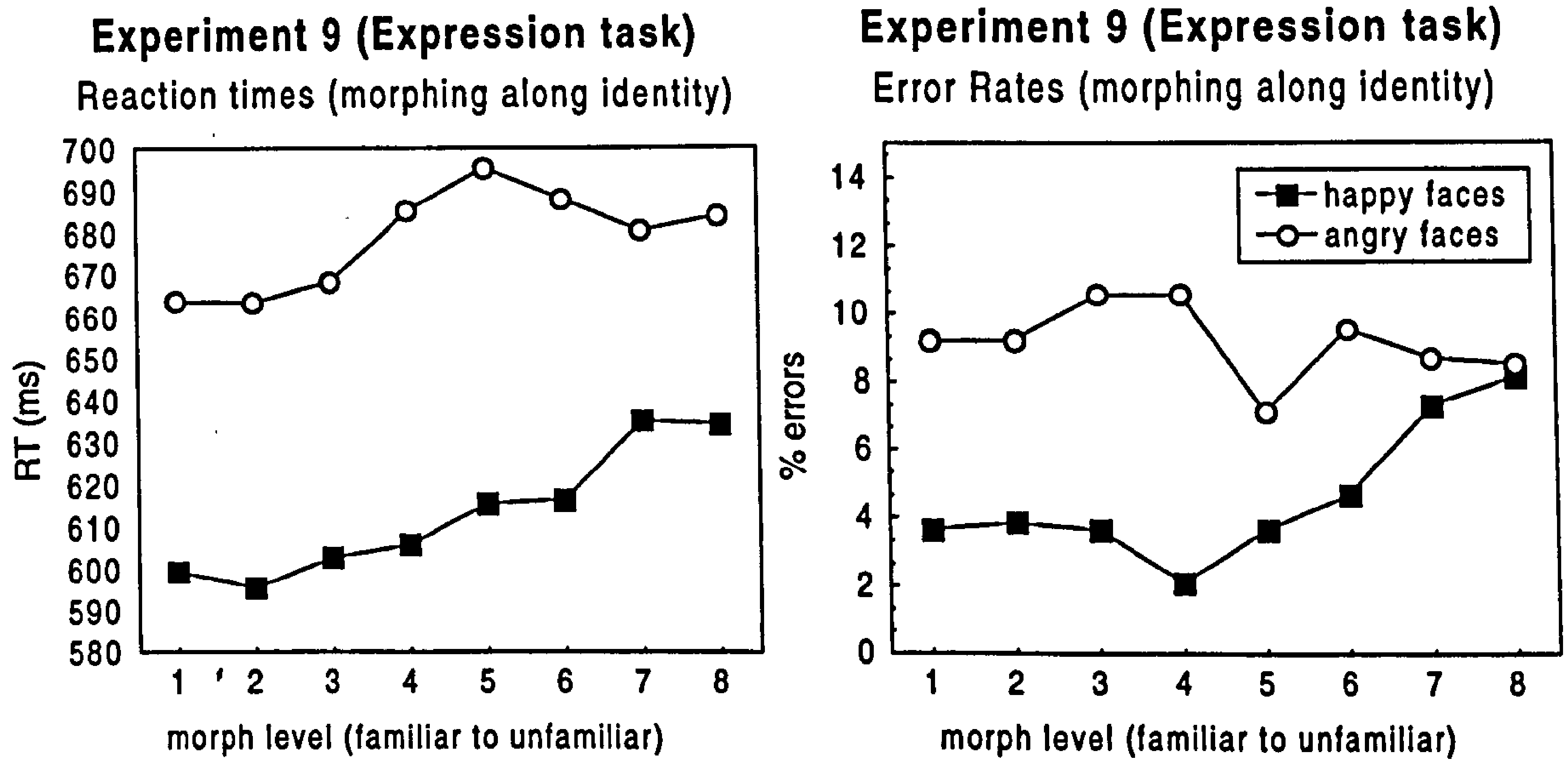

Figure 17: Results of Experiment 9 (classifications of emotional expression). RTs (left) and error rates (right) for faces morphed along identity within a constant expression. Square symbols demonstrate data for happy faces, circles depict the data for angry faces.

The analysis including both happy and angry faces revealed a highly significant linear trend, $F(1, 17) = 35.4, p < 0.001$, which suggested increasing RTs from familiar to unfamiliar faces (see also Figure 17). A quadratic trend for morph level interacted with expression, $F(1, 17) = 9.3, p < 0.01$.

A separate analysis of polynomial contrasts for happy faces revealed a highly significant linear trend, $F(1, 17) = 36.0, p < 0.001$, with increasing RTs from familiar to unfamiliar happy faces. For angry faces, in addition to a significant linear trend, $F(1, 17) = 8.8, p < 0.01$, which suggested increasing RTs from familiar to unfamiliar faces, there was also a quadratic trend, $F(1, 17) = 7.2, p < 0.05$, with maximum RTs for intermediate morph levels.

5.5.2.1.2 Error rates:

There was no main effect of morph level, $F(7, 119) = 1.7, p = 0.14$. Overall, error rates were higher for angry expressions, $F(1,17) = 9.2, p < 0.01$ ($M = 9.2\%$ vs. $M = 4.6\%$ for angry and happy expressions, respectively). This effect was further qualified by a significant interaction between expression and morph level, $F(7, 119) = 4.7, p< 0.001$ (see also Figure 17). The interaction was explored by separate ANOVAs for happy and angry faces, which revealed a highly significant effect of morphing from familiar to unfamiliar faces for happy expressions, $F(7, 119) = 7, p < 0.001$, while there was no morphing effect for angry faces $F(7, 119) = 1, p > 0.39$.

An analysis of polynomial contrasts including only happy expressions revealed both a linear, $F(1, 17) = 13.4, p < 0.01$, and a quadratic trend, $F(1, 17) = 20.3, p < 0.001$, suggesting overall increasing error rates from familiar to unfamiliar faces, with numerically lowest error rates for morph level 4. These analyses suggest that RT effects were not due to a speed accuracy trade-off.

5.5.2.2 Morphs along expression (identity constant):

5.5.2.2.1 Reaction times:

The ANOVA revealed a significant main effect of familiarity, $F(1, 17) = 18.6, p < 0.001$, reflecting faster expression classifications for familiar faces ($M = 676$ ms vs. $M = 694$ ms). As expected, RTs were also significantly influenced by the morphing procedure, *F*(7, 119) = 33.4, $p < 0.001$, demonstrating maximum RTs for the intermediate morph levels(see also Figure 18).

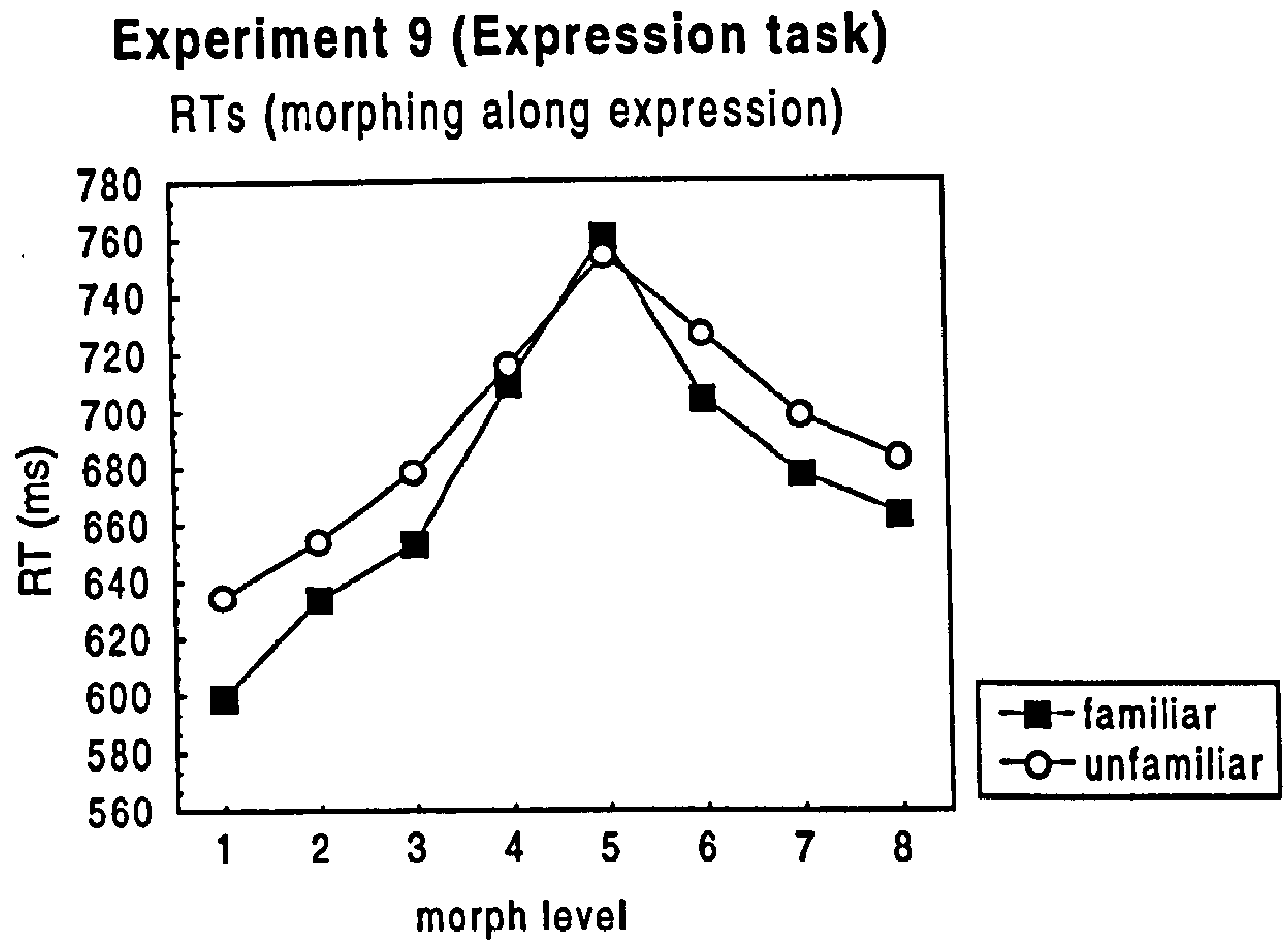

Figure 18: Results of Experiment 9 (classifications of facial expression). Classification response times for faces morphed from a happy to an angry expression within a given identity. Square symbols show data for familiar faces, circles depict the data for unfamiliar faces.

5.5.2.2.2 Classifications:

In addition to the expected highly significant main effect of morph level, *F*(1, 119) = 640.3, $p < 0.001$, which reflects relatively sharp category boundaries, there was a significant effect of familiarity on classifications of emotional expressions, *F*(1, 17) = 12.3, $p < 0.01$, suggesting that familiar faces were more likely to be classified as "happy" (57% vs. 53%, familiar and unfamiliar faces, respectively). This effect was further qualified by a significant interaction between familiarity and morph level *F*(7, 119) = 3.47, $p < 0.001$. Visual inspection of Figure 19 suggests that on the angry end of the expression morph continuum classifications were similar for familiar and unfamiliar faces, while happy and moderately happy faces seemed to evoke a higher percentage of "happy" responses for familiar faces.

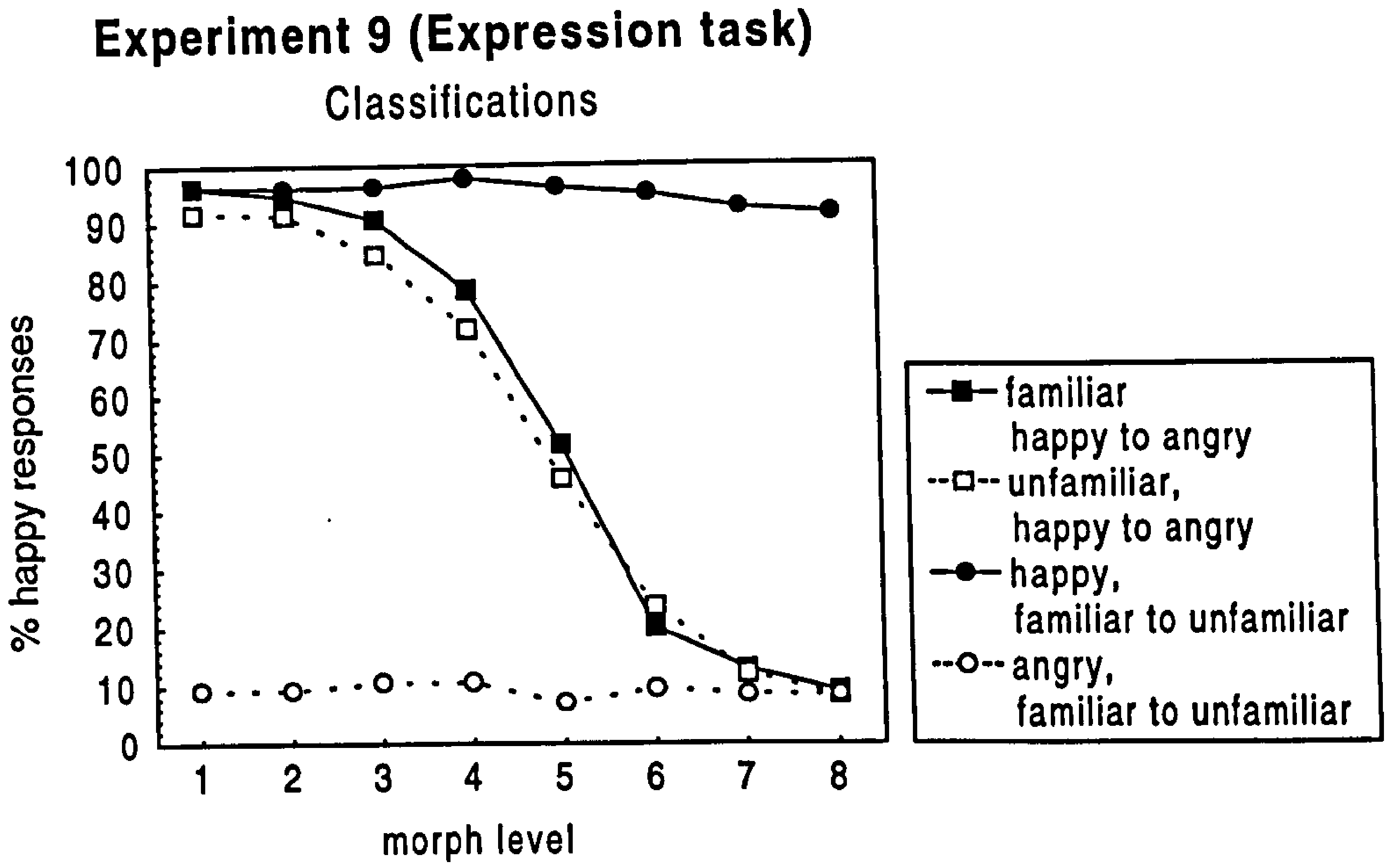

Figure 19: Results of Experiment 9 (classifications of emotional expressions). Percentage of "happy" classifications depending on morph level and familiarity. Square symbols correspond to data for faces morphed from a happy to an angry expression within a given identity; circles indicate data for faces morphed along identity for a constant emotional expression.

This impression was confirmed by separate ANOVAs comparing familiar and unfamiliar faces on each morph level. They revealed significant differences between familiar and unfamiliar faces for morph level 1, *F*(1, 17) = 12.36, $p < 001$, and for morph level 3, *F*(1, 17) = 19.7, $p < 0.001$ (morph level 1: $M = 96.4\%$ vs. $M = 91.8\%$ happy classifications for familiar and unfamiliar faces, respectively; morph level 3: M = 90.6% vs. $M = 84.7\%$ happy classifications for familiar and unfamiliar faces, respectively). Differences for morph level 2, *F*(1, 17) = 7, $p = 0.017$ and morph level 4, *F*(1,17) = 9.2, $p = 0.007$ were just marginally significant at the corrected α-level of 0.006 (morph level 2:

M = 94.8% vs. *M* = 91.3%; morph level 4: *M* = 78.5% vs. *M* = 71.8% happy classifications for familiar and unfamiliar faces, respectively).

5.5.3 Discussion

These data confirm previous findings of narrow category boundaries for classifications of emotional expressions (Schweinberger et al., 1999; Young et al., 1997). The finding that overall, happy emotional expressions were classified faster than other basic expressions has been reported before (Ekman et al., 1982; Kirouac et al., 1983). It has been suggested that this might be due to a more holistic processing for happy expressions in contrast to a more analytic processing of other basic emotions (Kirita et al., 1995).

In an identity classification task a strong correlation between familiarity and the identity categorization effect has been reported (Beale et al., 1995). The results in this expression categorization task suggest similar expression category boundaries for familiar and unfamiliar expression morphs. However, more consistent answers for familiar faces were found on the happy side of the expression continuum, while classifications of angry expressions did not differ between familiar and unfamiliar faces. In Experiment 8 faster and more accurate familiarity decisions for smiling familiar faces were found. This might provide further evidence that structural information at the level of FRUs (Bruce et al., 1986) might be stored together with typical expressions of familiar faces. Error rates for decisions on expression did not differ significantly between familiar and unfamiliar faces for angry displays.

Reaction times for expression classifications did not only demonstrate an effect of morphing along the expression dimension, but also increased along the familiar-unfamiliar continuum. In particular, there was a linear RT increase from familiar to unfamiliar faces for both expressions, while morphing along identity also showed a quadratic trend for angry faces. On the whole, error rates seemed basically to reflect the reaction time effects. Error rates for happy faces increased linearly along the identity continuum. However, according to the quadratic trend, error rates seemed to be smallest for the intermediate morphs. For angry expressions, error rates were not influenced by familiarity. The results extend those of Schweinberger et al. (1999), who reported no influences of morphing across two unfamiliar faces in an expression classification task using happy and angry faces. Here, performance for happy expressions decreased with increasing distance to the starting point of the identity morph continuum representing the original familiar faces. The findings might argue against a model of a strictly modular and independent processing of facial identity and emotional expression. The described effects might be due to an interaction between FRUs (Bruce et al., 1986) and processes that analyse facial expression. If the expression processing system was able to make use of idiosyncratic identity specific expressions, emotional expressions should be processed faster for highly familiar faces. At first sight, the present data are line with such an interpretation. It has to be noted that the results area at variance with experimental studies that did not find an influence of familiarity in expression matching tasks (Bobes et al., 2000; Bruce, 1986; Campbell et al., 1996a; Young, Hay & McWeeny, 1986). However, task requirements differ between expression matching and classification. Stored emotional "prototypes" might be more likely to influence decisions in speeded categorization than in matching, because directed visual processes might play a less significant role in classification tasks.

However, the present results have to be interpreted with caution because limited information is available about the "expressiveness baseline" for the familiar and unfamiliar stimulus sets. Although care has been taken to match the original familiar and unfamiliar faces that were used as basis for the morph stimuli as closely as possible, they still might have differed with respect to overall expressiveness. In addition, participants might have noticed a difference between the necessarily posed expressions displayed by unfamiliar faces, and the possibly more authentic expressions displayed by the familiar faces. It was therefore decided to test the familiar and unfamiliar picture sets for differences in a priori expressiveness by presenting them to participants who had no or at least a much lower degree of familiarity to the celebrities than the participants in Germany. A move to Scotland made it possible to re-run Experiment 9 at the University of Glasgow, Scotland, with British undergraduate students who had a significantly lower degree of familiarity to the celebrities. Similar RT increases from "familiar" to unfamiliar stimuli for participants who are unfamiliar with all faces would imply that differences in expressiveness between the original picture sets might have produced the identity morph effect on classifications of expression in Experiment 9.

5.6 EXPERIMENT 10

5.6.1 Method

5.6.1.1 Participants

The Experiment was conducted at the University of Glasgow, Scotland. In order to recruit participants, a poster was attached to a blackboard in the entry area of the Department of Psychology. It showed unedited portraits of all sixteen persons presented in Experiments 8 and 9. All photographs were different from the ones used in the experiment. The faces displayed a neutral expression and the original background was preserved. It was pointed out that subjects could only participate if they were unfamiliar with all of the shown persons.

In an attempt to ensure that all participants included in the analysis were unfamiliar with all faces, subjects were given a questionnaire after taking part in the experiment. It showed printouts of all faces, one portrait per page. For each face participants completed a 5-point rating scale, with "very familiar" coded as "4" and "never seen before" coded as "0". For any ratings other than "0" participants were asked to indicate the profession or name of the person. Only data of subjects who rated at least five out of the eight celebrities as completely unfamiliar (rating = 0) were included in the analysis. This criterion was reached by eighteen participants. (twelve women and six men) aged 19-31 years (*M* = 21.3 years, *SD* = 2.1 years; familiarity ratings: *M* = 0.29 and *M* = 0.09 for the celebrities and unfamiliar faces, respectively). Four participants correctly identified maximally one person. One subject was able to name two of the celebrities, while all others could not correctly identify any of the faces. All participants received 3.50 GBP (Pounds Sterling).

Data from thirteen additional participants were excluded, because they rated at least four out of the eight celebrities as familiar (rating > 0; *M* = 1.2). Data from two additional participants were excluded from the analysis because their error rates exceeded 25% in at least one experimental condition where emotional expression was not manipulated (Mean across participants: *M* = 4.8%).

5.6.1.2 Procedure

Stimuli, presentation and task were identical to Experiment 9.

5.6.2 Results

Overall, 96.5% of the original happy "familiar" and 95% of the original happy unfamiliar faces were classified correctly. For the original angry expressions, "familiar" faces yielded 87.8% and unfamiliar faces 92.2% correct responses. Missing answers were extremely rare ($M < 0.7\%$) and were not analysed further.

ANOVAs with identical repeated measurement factors as in Experiment 9 were performed. For comparisons between Experiment 9 and 10, ANOVAs including an additional between subjects variable "site" (Germany vs. Scotland) were performed.

5.6.2.1 Morphs along identity (expression constant):

5.6.2.1.1 Reaction times:

Overall, happy expressions were classified faster than angry expressions, $F(1, 17) = 109.0$, $p < 0.001$. As in Experiment 9, there was a significant main effect of morph level, $F(7, 119) = 5.8$, $p < 0.001$.

An additional ANOVA including data from Experiment 9 and 10 and site as between subjects factor showed no overall difference between both experiments, $F(2, 24) = 2.9$, $p > 0.1$. Notably, the significant interaction between site and morph level $F(7, 238) = 2.5$, $p < 0.05$, indicated that the effects of morphing along identity differed between the German and the Scottish sample.

However, an analysis of polynomial contrasts for Experiment 10 also revealed a significant linear trend, $F(1, 17) = 12.3$, $p < 0.01$, suggesting increasing RTs from "familiar" to unfamiliar faces (see Figure 20). In addition, there was a also quadratic trend, $F(1, 17) = 5.8$, $p < 0.05$.

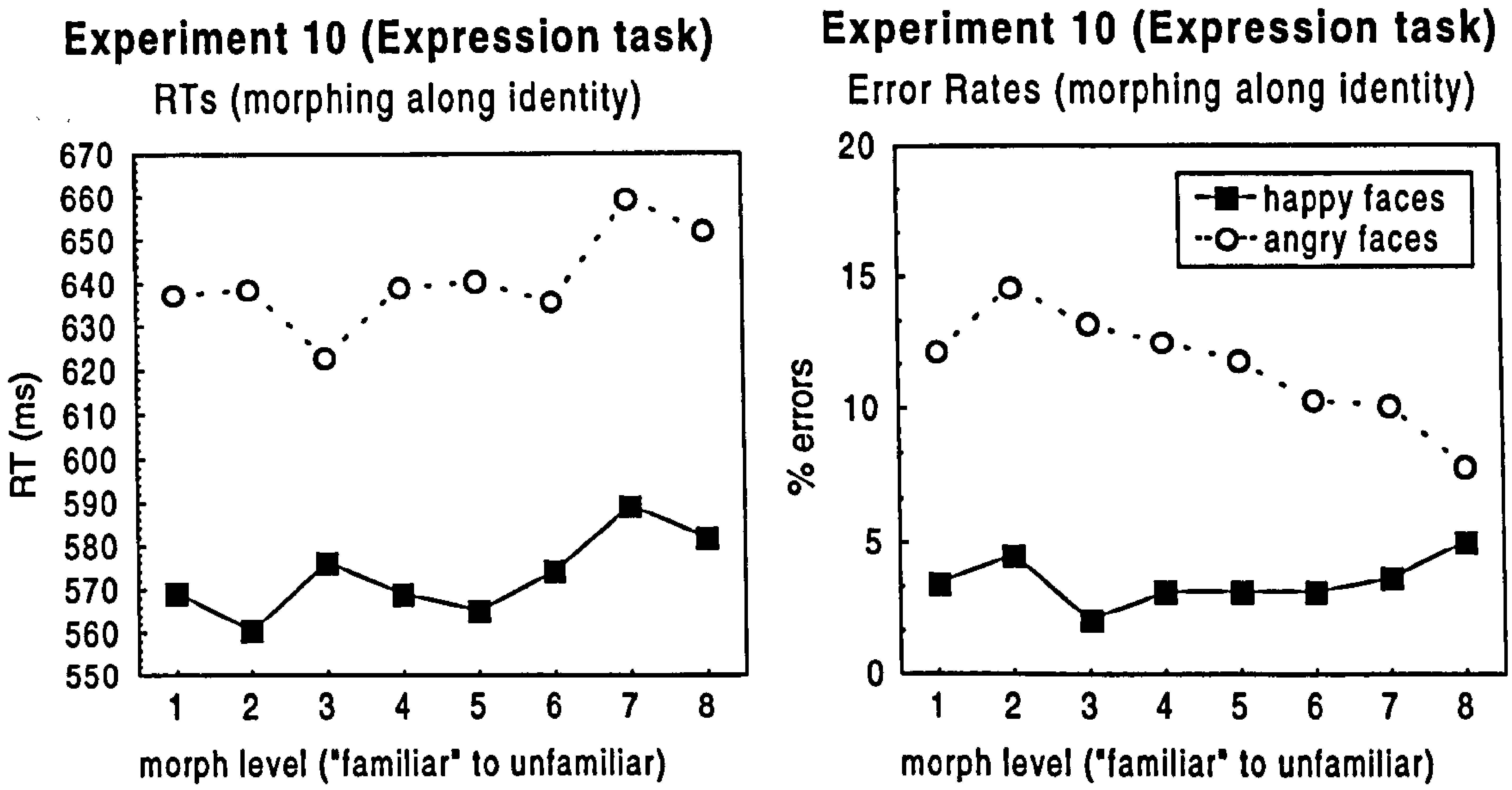

Figure 20: Results of Experiment 10. RTs (left) and error rates (right) for faces morphed along identity for a given emotional expression. Square symbols demonstrate data for happy, circles depict the data for angry faces.

A further analysis of polynomial contrasts including data from both experiments suggested that the linear trend was only marginally influenced by site, $F(1, 34) = 3.1$, $p < 0.09$, while the interaction was significant for the quadratic trend, $F(1, 34) = 5.9$, $p < 0.05$.

5.6.2.1.2 Error rates:

The analysis of errors of commission revealed a significant main effect of expression, suggesting overall lower error rates for happy faces, $F(1, 17) = 59.5$, $p < 0.001$, ($M = 3.5\%$ vs. $M = 11.6\%$, for happy and angry expressions, respectively). The effect of morphing from "familiar" to unfamiliar faces was significant, $F(7, 119) = 2.3$, $p < 0.05$. This effect was further specified by a highly significant interaction between expression and morph level, $F(7, 119) = 3.5$, $p < 0.01$, which was further explored by separate ANOVAs for happy and angry faces. For happy expressions, there was no effect of morphing along identity, $F(1, 17) = 1.7, p > 0.14$. while for angry expressions, error rates seemed to decrease from "familiar" to unfamiliar faces, $F(7, 119) = 3.5, p < 0.01$ (see also Figure 20).

An analysis of polynomial contrasts for angry expressions in Experiment 10 yielded a linear trend, $F(1, 17) = 5.3, p < 0.05$, suggesting decreasing error rates from "familiar" to unfamiliar faces. This implies that there might have been some speed-accuracy trade-off for angry identity morphs in Experiment 10.

An ANOVA including data from Experiment 9 and 10 yielded a highly significant interaction between site and morph level, $F(7, 238) = 3.0, p < 0.01$. This interaction was not further specified by expression, as demonstrated by the non-significant three-way interaction between site, morph level and expression, $F < 1$.

5.6.2.2 Morphs along expression (identity constant):

5.6.2.2.1 Reaction times:

The ANOVA indicated shorter RTs for "familiar" compared to unfamiliar faces, $F(1, 17) = 13.9, p < 0.01$ ($M = 637$ ms vs. $M = 650$ ms). The expected morph level main effect, $F(7, 119) = 66.2, p < 0.001$ interacted only marginally with "familiarity", $F(7, 119) = 2.18, p = 0.06$ (see Figure 21).

An ANOVA including data from Experiments 9 and 10 did not show an interaction between site and "familiarity", $F(1, 34) < 1$, suggesting that across all morph levels, RT differences between "familiar" and unfamiliar faces were of similar size in both experiments ($Mdiff = 18$ ms vs. $Mdiff = 13$ ms in Experiments 9 and 10, respectively).

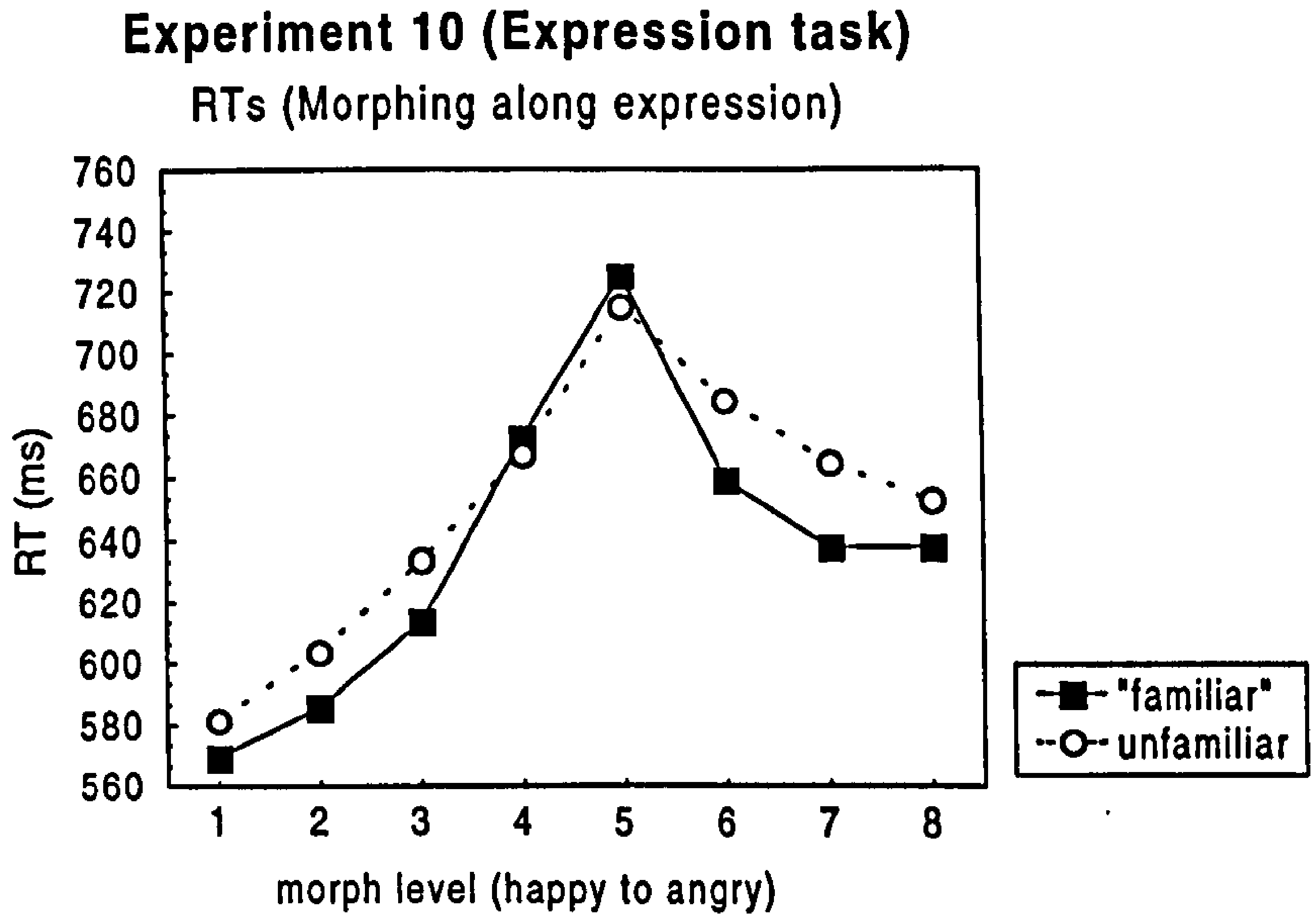

Figure 21: RTs for faces morphed between expressions within a constant identity in Experiment 10.

5.6.2.2.2 Classifications:

There was no overall effect of "familiarity", $F(1, 17) < 1$. The expected effect of morph level, $F(7, 119) = 521.1$ $p < 0.001$, only interacted marginally with "familiarity", $F(7, 119) = 2.11, p = 0.065$ (see also Figure 22).

An ANOVA including data from the German and the Scottish sample revealed a significant interaction between "familiarity" and morph level, $F(7, 238) = 3.5, p < 0.01$, and a marginally significant three-way interaction between site, "familiarity" and morph level, $F(7, 238) = 2.1, p = 0.059$.

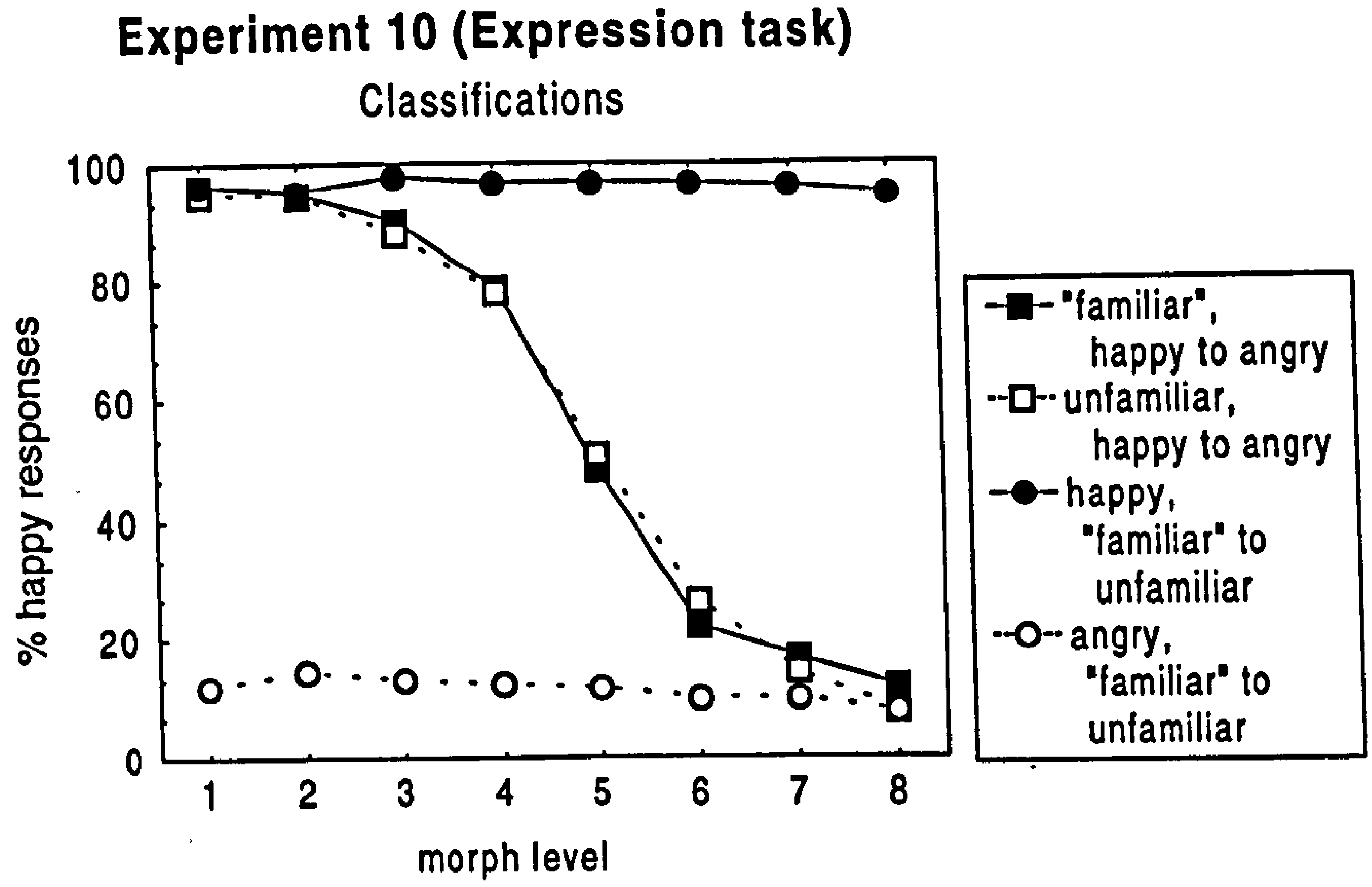

Figure 22: Percentage of "happy" classifications depending on morph level and "familiarity" in Experiment 10. Square symbols show expression morphs, circles depict data for morphs between identities.

5.6.3 Discussion

Overall, the results in Experiment 10 are similar to the ones in Experiment 9. Importantly, there was a linear increase of RTs for morphs from "familiar" to unfamiliar faces in both experiments and this linear trend interacted only marginally with experimental site. This suggests that the influences of familiarity on classifications of emotional expressions found in Experiment 9 might have been at least partially due to differences in general expressiveness between the familiar and the unfamiliar original picture sets.

However, participants in Germany (Experiment 9) had demonstrated a "happy-bias" for familiar faces: familiar faces were more likely to be classified as "happy". This was not the case in Scotland. Furthermore, the interaction between familiarity and expression morph level found in Germany showed that the first four morph levels along the expression continuum were more likely to be classified as "happy", when they were displayed by familiar faces. Pictures on the “angry half” of the morph continuum were classified in a similar way for familiar and unfamiliar faces. This interaction between familiarity and expression morph level was influenced by the factor “site”. In Germany, participants were more likely to classify expressions as happy, if the poser was familiar. In Scotland, where all faces were unfamiliar, this bias was not observed. However, morphs along emotional expressions were classified faster for “familiar” faces, and this effect did not interact significantly with site. A similar picture emerged for morphs along identity. Overall, morphing from familiar to unfamiliar faces had produced increasing RTs in Germany. Although this morph level main effect interacted significantly with site, the linear trend, which was observed in both experiments only interacted marginally with site.

However, some evidence for an influence of familiarity comes from the analysis of error rates. In Germany, error rates had only increased systematically along the identity continuum for happy expressions. In Scotland, error rates for happy faces were not influenced by “familiarity”. For angry faces, error rates linearly decreased from “familiar“ to unfamiliar faces. Here, participants in Scotland might have traded speed for accuracy, which might partly explain longer RTs for unfamiliar faces.

Finally, the influence of morphing along identity interacted with site, suggesting qualitative differences between the two Experiments. Importantly however, this did not yield for the linear trend of RTs of expression classifications for morphs from “familiar” to unfamiliar faces.

To sum up, the findings of an facilitating effect of familiarity on the analysis of expression in Experiment 9 has to be interpreted with caution. Although there might be an effect of familiarity in the sense that participants in Experiment 9 tended to associate familiar faces with happy expressions, which is reflected by a "happy bias" for familiar faces, the finding that familiarity facilitates the processing of facial expression was not substantiated unequivocally in Experiment 10. Thus, the possibility remains that the advantage for processing expressions from familiar relative to unfamiliar faces (as seen in Experiment 9) was at least partially due to differences in expressiveness between the pictures used in the familiar and unfamiliar sets. However, it has to be mentioned that this is difficult to decide at present, because there might have been residual differences in familiarity between the sets in Experiment 10, as the familiarity ratings were not equal to zero. This issue therefore requires more stringent investigation.

5.7 General Discussion

Experiments 8-10 investigated dependencies between the processing of facial identity and emotional expression using morphed stimuli in speeded two-choice classification tasks. The experiments were designed with respect to a model of face perception, which assumes modular and independent parallel processing of identity and emotional expression (Bruce and Young, 1986). The model predicts that face recognition is independent of systematic variations of emotional expressions and that classifications of emotional expressions are not modulated by familiarity.

In Experiment 8, morphing along emotional expressions influenced classification speed and error rates only for familiar faces, while no modulation was observed for unfamiliar faces. While the results replicate and extend findings of sharp category boundaries for the perception for identity (Beale et al., 1995; Schweinberger et al., 1999), they cannot be accounted for by a model which assumes completely expression independent representations of familiar faces. Also, the notion of a "normalization" of emotional expressions when classifying facial identity is not in line with the present results.

Data on the relationship between face recognition and expression analysis including a range of various emotions are extremely sparse in the literature. A number of experimental studies did not find influences of emotional expression on the processing of identity. Most of these studies used matching paradigms (Bobes et al., 2000; Campbell et al., 1996a; Young et al., 1986) which might not provide an optimal approach to explore stored representations of faces because face matching might require a higher degree of directed visual processing than the classification task applied here and might underestimate the influence of person specific visual memory. Furthermore, to my knowledge no study so far systematically manipulated expressions of familiar faces the way it was done here by means of morphing. In most studies, the null-hypotheses was accepted and generalized across the processing of expression in general, although only stimuli displaying a limited number of expressions were used.

However, there seems to be evidence for facilitated face recognition by certain types of expressions (Davies et al., 1982; Endo et al., 1992; Kottoor, 1989; Sansone et al., 1996; all cited in: Baudouin et al., 2000). Endo et al. (1992) found better recognition of famous faces when they were smiling, whereas students recognized their teachers better when pictures displayed neutral expressions. In a recent study, smiling faces were more likely to be classified as familiar than neutral faces (Baudouin et al., 2000). These results are in line with the present data and suggest that under certain conditions face recognition might be influenced by emotional expression. An explanation of the results in Experiment 8 in terms of superficial pictorial information (Bruce, 1982) available only for pictures of celebrities can be ruled out. It is particularly striking that face recognition was fastest and most accurate for the artificial morph stimuli that had never been seen by the subjects before taking part in the experiment.

A parallel model of identity and expression, which allows for top-down processing via the cognitive system and Person Identity Nodes (Burton et al., 1990) might explain more efficient face recognition for expressions that are highly associated with familiar persons via top-down semantic activation. However, as this would include some additional processing time for a cognitive component following the analysis of expression, such an explanation would require slower response times in the identity compared to the expression task due to the additional component. This was not the case for the used stimuli: an ANOVA comparing Experiment 8 (identity task) and Experiment 9 (expression task) across all morph levels revealed shorter RTs for the identity task, $F(1, 33) = 12{,}3$, $p < 0.01$ ($M = 598$ ms vs. $M = 668$ ms for Experiment 8 and 9, respectively).

An alternative explanation which still allows for a parallel processing of both dimensions might be that structural information used to recognize familiar faces might preserve some information about typical expressions, possibly creating an "emotional prototype" for each familiar person. A prototype effect refers to a tendency to recognize a face corresponding to the central value of a series of seen faces, even when this central value or prototype has not been seen (e.g. Cabeza et al., 1999). Possibly, the construction of such a prototype might include information about person specific "typical" expressions. At the moment it can only be speculated whether better recognition for smiling familiar faces originated from stored information about "typical" expressions, because information about the frequency with which participants had encountered an expression displayed by a particular face during familiarization is not available for the faces used here. However, an inspection of journals and other media suggests that most celebrities are more likely to be seen with smiling than angry expressions. More research using controlled exposure to emotional displays during face learning is necessary to answer this question.

It has previously been shown that both for the processing of identity and expression, configural information is extracted from a face (Calder et al., 2000). If the face recognition system was able to take idiosyncratic configurational information into account when analysing expressions, performance should be better for expressions displayed by familiar faces. Experiments 9 and 10 demonstrated an influence of familiarity in the sense of a higher probability to associate familiar faces with happy expressions. However, there was no unequivocal effect for classification response times in the expression tasks. The finding of linearly increasing RTs with decreasing familiarity for participants in Germany (Experiment 9) was also observed in a control study including British participants who were unfamiliar with all faces (Experiment 10). The linear trend only marginally interacted with site. This implies that differences between the familiar and unfamiliar faces may have been at least partially due to differences between the stimulus sets with respect to expressiveness, rather than representing an influence of familiarity on the processing of expression. This would be in line with a number of studies that did not find an influence of familiarity on expression matching (Bobes et al., 2000; Bruce, 1986; Campbell et al., 1996a; Young et al., 1986). However, there are some data supporting the notion of contingencies between identity and expression processing from studies that used speeded classification instead of matching tasks. Baudouin et al. (2000) reported that expressions displayed by familiar faces could be recognized better than expressions of unfamiliar faces when the mouth was covered or presentation time was short. There is a report that speeded classifications of emotions from video clips were influenced by familiarity (Peng, 1989; cited in: Campbell et al., 1996a).

To summarize, there was some evidence for an influence of familiarity on expression classification in the sense that familiar faces were more likely to be classified as happy. However, the experiments do not provide clear-cut evidence of a faster expression processing for familiar faces. The findings demonstrate an influence of expression on face recognition, while classifications of unfamiliar faces were unaffected by morphing along expression. It was argued that this pattern might be caused by expression dependent representations at the FRU level. Future research has to demonstrate whether similar effects can be found with morph stimuli of experimentally familiarized faces which would allow for a controlled exposure to specific identity-expression combinations.

6 Experiments 11 and 12: Influence of relative processing speed of identity and expression in the selective attention paradigm. Experiments with morphed faces

6.1 Purpose of Experiments 11 and 12

The experiments aimed at further investigating the relationship between the processing of identity and emotional expression by applying Garner's paradigm of selective attention. A recent study that had used this paradigm (Schweinberger et al., 1998) suggested an "asymmetric relationship" between the processing of facial identity and emotional expression in the sense that identity is processed independently of expression but not vice versa. In contrast, Experiment 6 and 7 did not show an asymmetric relationship between these two dimensions. It was argued that three possible explanations could have accounted for the conflicting results. First, the lack of orthogonal interference in Experiment 6 and Experiment 7 might have been due to a successful control of a potentially asymmetric increase of task difficulty, which might have been present in the study by Schweinberger et al. (1998). The design of Experiment 6 and 7 controlled for differences in overall stimulus variability between blocks and made sure that the increase of task-relevant information from the control to the orthogonal condition was comparable in both tasks.

Alternatively, the diverging results might have been due to differences in the definitions of identity variations. In the experiments by Schweinberger et al. (1998), identity was defined with respect to *individual face identity,* while in Experiments 6 and 7, the identity dimension was defined in terms of a super-ordinate *face familiarity* *category*. Although the familiarity category was held constant in the control condition of the expression task, a variety of individual faces was presented. This might have produced orthogonal interference in the control condition and possibly levelled potential differences between the control and the orthogonal condition in the expression task.

Finally, another important aspect that differed between the study by Schweinberger et al. (1998) and the present Experiments 6 and 7 was overall task difficulty. Task difficulty as reflected by processing speed might differ between the identity and the expression dimension (Campbell et al., 1996a; Schweinberger et al., 1999), although this is probably dependent on the particular stimuli that are used. In fact, there was a trend for faster identity classifications compared to classifications of expression for the faces used by Schweinberger et al. (1998), while no such difference was found for the stimulus set presented in Experiments 6 and 7. Interestingly, Schweinberger et al. (1998) found an influence of identity variations on expression classifications, while classifications of identity were unaffected by variations of expressions. In contrast, no orthogonal interference in any of both tasks was found in Experiments 6 and 7.

Importantly, in the selective attention paradigm the more difficult dimension might be more affected by irrelevant variations of the easier dimension than vice versa. If this were the case, the findings of an asymmetric interaction between the processing of identity and expression (Schweinberger et al., 1998) might be at least partially stimulus dependent and might not be generalized. Schweinberger et al. (1999) investigated the role of perceptual saliency and relative processing speed for the asymmetric relationship between identity and expression processing. By means of morphing they selectively manipulated stimulus saliency either for the identity or the expression dimension and found that perceptual saliency had no influence on the previously reported asymmetric interaction in the selective attention paradigm. However, for the same reasons already outlined above (for a discussion see sections 2.5 and 4.1) the design of the study does not rule out the possibility that the results were produced by picture based response strategies or by an asymmetric increase of relevant information in the expression and the identity task. Therefore, the possibility remains that findings of an asymmetric interaction were mainly caused by differences in overall task difficulty and that the absence of orthogonal interference in Experiments 6 and 7 reflects similar task demands.

Experiments 11 and 12 addressed the question whether the absence of an asymmetric interaction between identity and expression classifications in Experiments 6 and 7 was due to identical processing speed in both tasks. This was tested by using a similar design as in Experiments 6 and 7, while the difficulty of the particular relevant dimension was selectively manipulated by means of morphing.

6.2 Rationale of Experiments 11 and 12

If the asymmetric interaction between the processing of identity and facial expression in the selective attention paradigm (Schweinberger et al., 1998) really reflects a stable functional architecture of face perception, the absence of such an interaction in Experiments 6 and 7 might have been due to differences with respect to the definition of identity variations in both studies. Alternatively, the reported asymmetric interaction might have been a product of overall differences in task difficulty and relative processing speed. Such an explanation might account for the absence of any orthogonal interference in Experiments 6 and 7, where both tasks yielded similar reaction times.

It was the aim of Experiments 11 and 12 to further investigate the conflicting results of Experiments 6 and 7 and the studies by Schweinberger et al. (1998; 1999). The overall design is similar to Experiments 6 and 7, but the crucial difference is that in Experiments 11 and 12, task difficulty will be selectively manipulated by means of morphing. The objective is to test whether orthogonal interference can be found, if the respective relevant dimension is more difficult to process than the irrelevant dimension. If this were the case, it would suggest that the absence of orthogonal interference in Experiments 6 and 7 was due to the fact that processing speed for the two tasks did not differ significantly for the used stimuli. Such a finding would limit the extent to which Schweinberger et al.'s (1998; 1999) results can be generalized.

It has been shown that morphing can be used to selectively manipulate either facial identity or facial expression information, leaving the respective other dimension relatively unaffected (Beale et al., 1995; Calder et al., 1996; Schweinberger et al., 1999; but see also Experiment 8). In the present experiments, the selective attention paradigm was applied to stimuli that were either easy or difficult to classify with respect to a particular task-relevant dimension, which was either expression (Experiment 11) or identity (Experiment 12). The task-irrelevant dimension was always highly salient and therefore easy to classify. The rationale was borrowed from Schweinberger et al. (1999), who used the starting and end points of morph continua between either two facial identities displaying the same expression or two emotional expressions within the same identity as highly salient (or easy) stimuli. Morph pictures on levels 3 and 6 on the respective task-relevant morph continuum, which were still classified consistently, but at lower speed (as also shown in Experiments 8 and 9) were used as less salient (or difficult) stimuli.

The design of the experiments was similar to Experiments 6 and 7. Participants performed speeded classifications of either identity or facial expression in a selective attention paradigm. As in Experiments 6 and 7, identity was defined in terms of a *face familiarity category* and there were variations with respect to *individual face identity* within familiarity categories and therefore within the control condition of the expression task. Similar to Experiments 6 and 7 pictorial response strategies and asymmetric increases of task-relevant information were controlled for by using a significantly larger stimulus set than in recent studies (Schweinberger et al. 1998; Schweinberger et al., 1999). Because of the limited informative value of the correlated condition and the potential problems regarding its interpretation, it was decided only to use control and orthogonal conditions in Experiments 11 and 12.

The following outcomes can be anticipated: an orthogonal interference of the relevant on the irrelevant dimension is only found, when the relevant dimension is significantly more difficult to classify than the irrelevant one. Such a result would argue for a crucial influence of relative processing speed on the reported asymmetric interaction (Schweinberger et al., 1998) and explain the absence of such an interaction in Experiments 6 and 7, where reaction times were the same in both tasks. Such a finding would suggest that reports of an asymmetric interaction between the processing of identity and expression do not generalize across different stimulus characteristics and probably do not reflect a fixed architecture of face perception processes.

Alternatively, irrespective of stimulus saliency, orthogonal interference might be found neither in the identity nor in the expression task. In this case the question whether conflicting results between Experiments 6 and 7 and the studies by Schweinberger et al. (1998; 1999) were either caused by different definitions of identity variations or were due to an efficient control of an asymmetric increase of task-relevant information from the control to the orthogonal conditions in Experiment 6 and 7 would have to be further investigated.

6.3 Method

6.3.1 Stimuli and Apparatus

A subset of the stimuli from Experiments 8 and 9 was used. These included the original photographs (morph levels 1 and 8) and morph levels 3 and 6 for expression and identity morphs. Experiments 8 and 9 had shown that faces morphed along expressions or identities are perceived in a relatively categorical manner, while increasing RTs reflect differences in stimulus saliency. Similarly to the study by Schweinberger et al. (1999), morph levels 3 and 6 were selected, because for these morph levels classifications were still consistent but significantly slower, suggesting a decrease of perceptual saliency or in other words an increase of relative task difficulty.

In an attempt to obtain a homogenous stimulus set that was classified consistently both with respect to expression and familiarity, six of the eight face pairs used in Experiments 8 and 9 were selected. Two face pairs were excluded because error rates for at least one of the difficult morph levels were too high (difference between error rates for a particular face pair and average error rates exceeded 20%).

In Experiment 11 (identity task), morph levels 1, 3, 6 and 8 of the identity morphs as described for Experiment 8 were used. For morph level 1 and 8 identity was easy, and for morph levels 3 and 6 identity was difficult to classify. Faces either displayed a happy or an angry expression and the expression information was always highly salient (for examples of the stimuli see also Figure 23).

This resulted in a total stimulus set of 48 pictures (two easy familiar faces, happy and angry; two easy unfamiliar faces, happy and angry; two difficult familiar faces, happy and angry; two difficult unfamiliar faces, happy and angry, for six face pairs each:).

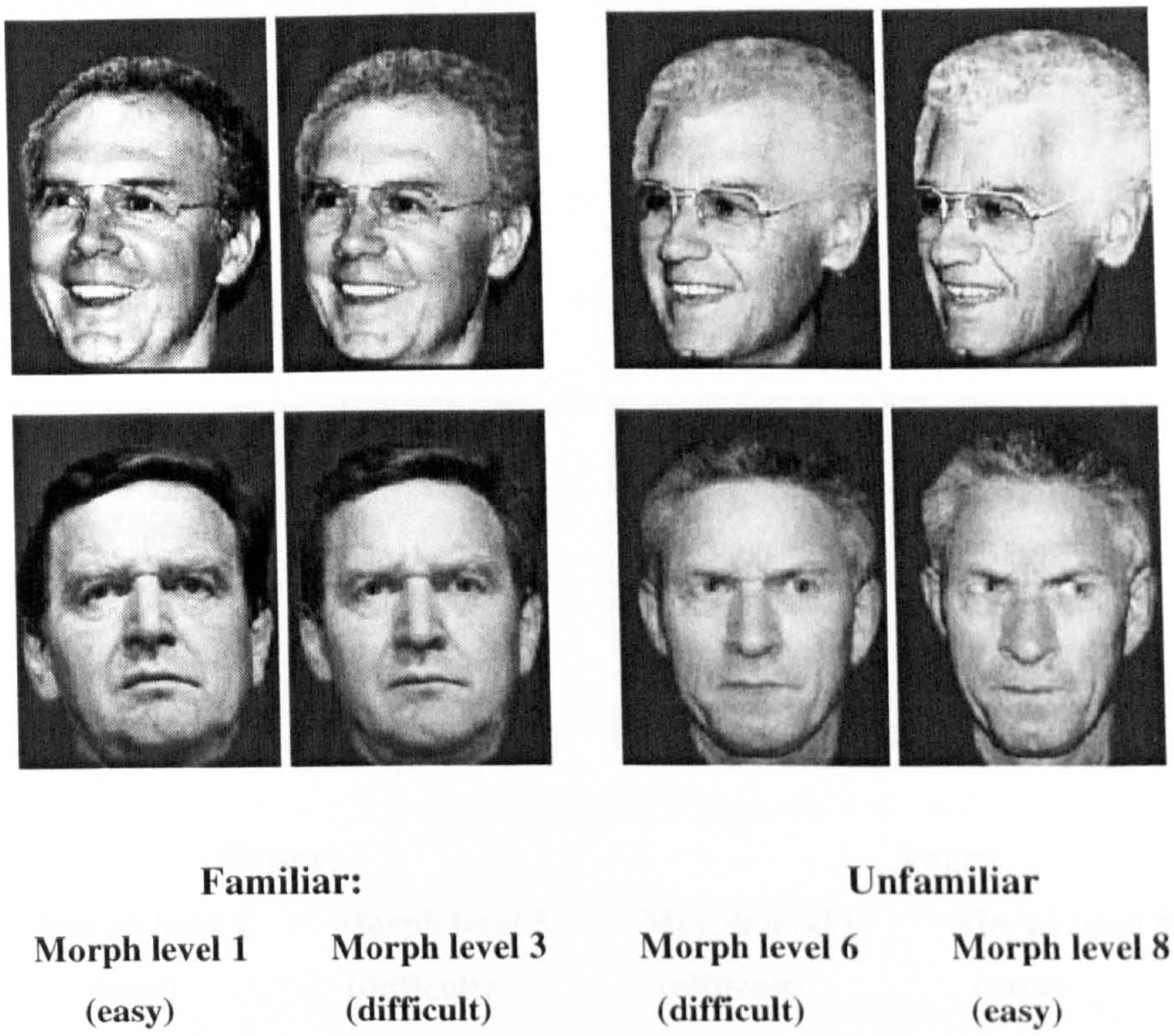

Figure 23: Examples of stimuli used in Experiment 11. Top row: morphs from a familiar to an unfamiliar face, within a happy expression. Bottom row: morphs from a familiar to an unfamiliar face within an angry expression (morph levels 1, 3, 6, 8 from left to right).

In Experiment 12, morph levels 1, 3, 6 and 8 of the expression morphs were used. Accordingly, for morph levels 1 and 8, the expression dimension was easy, and for levels 3 and 6 difficult to classify. Faces were either familiar or unfamiliar and the identity dimension was highly salient for all stimuli (for examples see also Figure 24).

This resulted in a total stimulus set of 48 pictures (two easy happy faces, familiar and unfamiliar; two easy angry faces, familiar and unfamiliar; two difficult happy faces, familiar and unfamiliar; two difficult angry faces, familiar and unfamiliar, for six face pairs each).

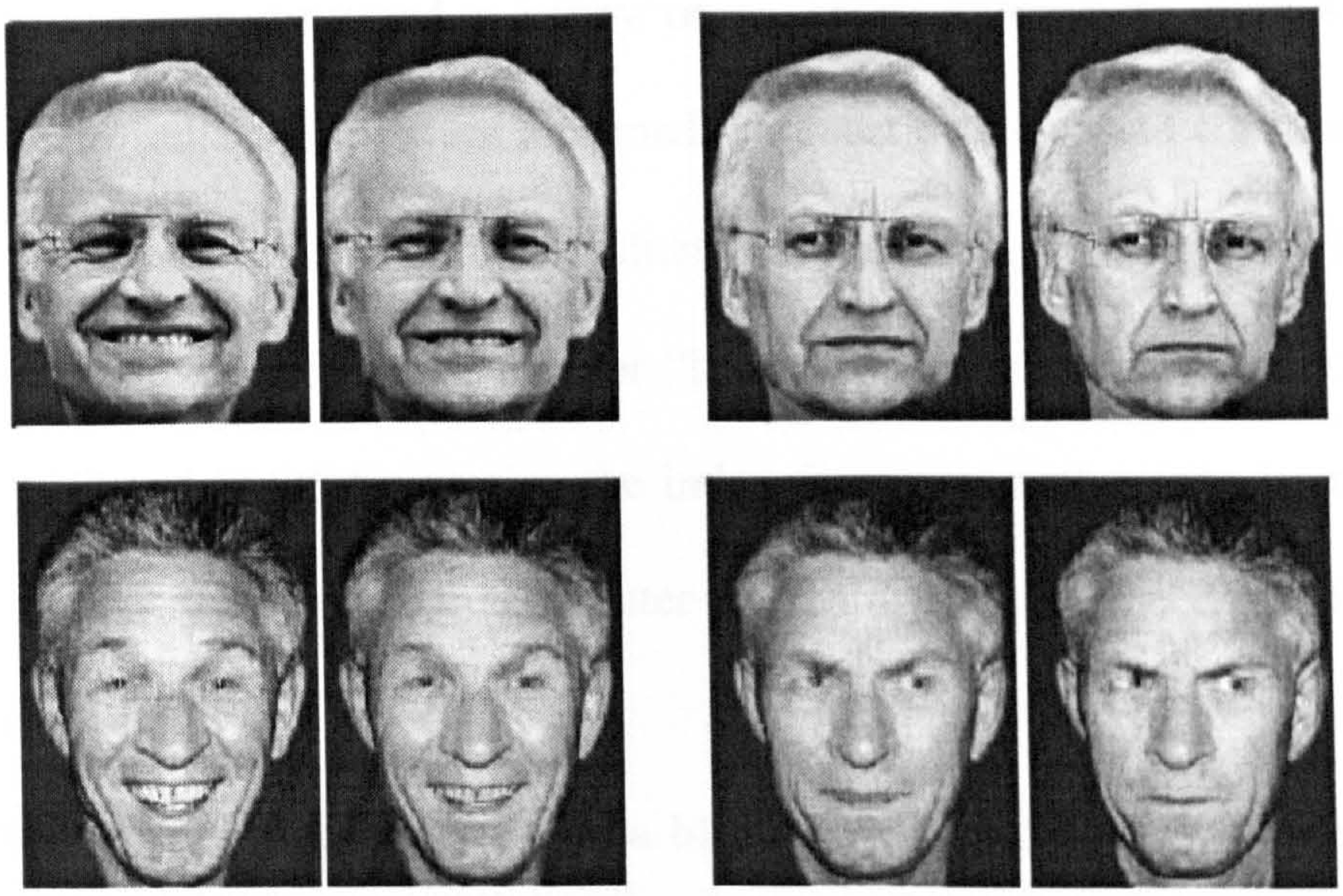

Figure 24: Examples of stimuli in Experiment 12. Top row: morphs from a happy to an angry expression for a familiar face. Bottom row: morphs from a happy to an angry expression for an unfamiliar face (morph levels 1, 3, 6, 8 from left to right).

The stimuli were presented on black background in the centre of a 19'' monitor that was connected to an IBM compatible personal computer. The presentation software was ERTS^{TM} (Experimental Runtime System, Berisoft Corporation). Picture resolution was 17.7 pixels/cm at a screen resolution of 800 by 600 pixels The size of the stimuli was 10 cm x 7.5 cm at a viewing distance of 60 cm, resulting in a vertical visual angle of 9.5 degrees and a horizontal visual angle of 7.1 degrees.

6.3.2 Procedure

In Experiment 11, participants discriminated in a speeded two-choice task whether the face was either familiar or unfamiliar. In Experiment 12, participants decided, whether the face either displayed a happy or an angry expression. Both speed and accuracy were stressed. Subjects were informed that reaction times and error rates would be recorded. Responses were made by simultaneously pressing the "F1" and "F12" keys using the middle fingers of both hands for "happy" or "familiar" and simultaneously pressing the "F2" and F"11" keys using the index fingers of both hands for "angry" or "unfamiliar" responses on a standard computer keyboard that was turned for 180 degrees for practical reasons.

In all trials a white fixation cross on a black background was shown for 500 ms, followed by a face stimulus visible for 2000 milliseconds or until a key was pressed. After a key-press, the face was replaced by a fixation cross for 1000 ms. Visual feedback in form of the words "too fast" or "too slow" (in German), presented for 500 ms was only given for fast (reaction times < 100 ms), missing and slow answers (reaction times > 1600 ms). After reading the instructions on the monitor, participants were shown examples of additional face stimuli that were not used in the experiments.

Each experiment consisted of four blocks. Two blocks formed the control and two the orthogonal condition. Within each block, the stimuli were presented in random order. Each block was followed by a break that lasted for at least twenty seconds. The end of the break was self-paced. The duration of the experiment was about twenty minutes.

In Experiment 11 (identity task), one block of the control condition, contained only happy faces of familiar and unfamiliar individuals. In the other block of the control condition, only angry familiar and unfamiliar faces were presented. All stimuli in each block of the control condition were repeated four times. In the orthogonal condition both familiar and unfamiliar faces displaying both expressions were shown. The orthogonal block was presented twice with a short break between both presentations. All stimuli in each block of the orthogonal condition were repeated twice.

In Experiment 12 (expression task), one block of the control condition showed only familiar faces, displaying happy and angry expressions. In the other block of the control condition, only unfamiliar faces displaying happy and angry expressions were presented. Stimuli were repeated four times per block. In the orthogonal condition both familiar and unfamiliar faces displaying both expressions were shown. The identical orthogonal set was presented twice with a short break between both presentations and stimuli were repeated twice per block. This procedure ensured that the same stimuli entered the analysis of the control and orthogonal condition per Experiments, ruling out that possible differences between both conditions might be due to stimulus inherent differences.

In both experiments all blocks were preceded by sixteen practice trials in order to familiarize participants with the respective experimental condition. Practice trials consisted of pictures from an additional face pair and were not analysed. The order of experimental conditions as well as the order of blocks within the control condition was completely counterbalanced across participants. The order of stimuli within blocks was random.

After the experiment, participants rated, how familiar they were with the celebrities by completing a 7-point rating scale, with "never seen before" post-hoc coded as "0" and "very familiar" post-hoc coded as "6". Only data of participants who achieved average ratings of at least $M = 2.5$ for the familiar faces were included. The average ratings for the familiar faces were $M = 5.3$ ($SD = 0.5$) in Experiment 11 and $M = 5.2$ ($SD = 0.5$) in

Experiment 12. In order to ensure that all participants were unfamiliar with all of the supposedly unknown faces, they were presented with one photograph of each unfamiliar face after the experiment and asked whether they knew the respective person. None of the participants had seen any of the unfamiliar faces before.

Data were averaged across face pairs. First, ANOVAs on RTs and error rates including data from both experiments were performed. In addition to the between-subjects factor experiment (identity task vs. expression task), four repeated measurement factors were included. These factors were condition (control vs. orthogonal), task-relevant dimension (familiar vs. unfamiliar in the identity task and happy vs. angry in the expression task), task-irrelevant dimension (happy vs. angry in the identity task and familiar vs. unfamiliar in the expression task) and stimulus difficulty (easy vs. difficult). Because the particular alternatives of the relevant and irrelevant dimensions cannot be meaningfully compared across experiments, main effects or interactions including these factors are not reported for the overall analyses but were explored with separate ANOVAs per experiment.

For the analysis of RTs and error rates in Experiment 11, ANOVAS with repeated measurement factors on condition (control vs. orthogonal), expression (happy vs. angry), familiarity (familiar vs. unfamiliar) and difficulty (easy vs. difficult) were performed.

Accordingly, the ANOVAs for Experiment 12 included the repeated measurement factors condition (control vs. orthogonal), familiarity (familiar vs. unfamiliar), expression (happy vs. angry), and difficulty (easy vs. difficult).

When performing ANOVAs, α–levels for post-hoc ANOVAs were Bonferroni corrected. Only correct answers between 150 ms and 1500 ms were entered into RTs analyses.

6.3.3 Participants

Twelve participants (seven women and five men) aged 21 – 39 years (*M* = 25.8 years, *SD* = 5.7 years) contributed data in Experiment 11. The Experiment was conducted at the University of Konstanz, Germany. Participants received either a fee of 7.50 deutsche marks (DM; *n* = 10) or course credit (*n* = 2). Data from two additional participants were excluded from the analysis due to excessive error rates (*M* > 15%, Mean across participants: *M* = 5%).

Twelve different participants (seven women and five men) aged 20 – 28 years (*M* = 24.8 years, *SD* = 2.9 years) contributed data in Experiment 12. The Experiment was conducted at the University of Konstanz, Germany. Participants received either a fee of 7.50 deutsche marks (DM; *n* = 11) or course credit (*n* = 1). Data from one additional participant had been replaced due to a low familiarity score (*M* = 1.8). Data from two additional participants were replaced due to excessive overall error rates (*M* > 15%, Mean across participants: *M* = 6.9%). Data from one additional participant was replaced due to excessive error rates in at least one condition (*M* > 40%, Mean across participants for the respective condition: *M* = 12.6%).

6.4 Results

Missing, invalid and inconsistent answers (e.g. "F1" and "F11" key presses) were extremely rare both in Experiment 11 (*M* < 0.1%, *M* < 0.3% and *M* < 0.7%, respectively) and Experiment 12 (*M* < 0.2%, *M* < 0.9% and *M* < 0.6%, respectively) and were not analysed further

6.4.1 Reaction times

Overall, there was no effect of condition, $F(1, 22) < 1$, and no interaction between experiment and condition, $F(1, 22) < 1$. There was also no significant interaction between condition and difficulty, F(1, 22) < 1. However, the significant three-way interaction between experiment, condition and difficulty, $F(1, 22) = 4.7$, $p < 0.05$ reached significance (see also Figure 25). Identity was classified faster than expression, as suggested by the significant main effect of experiment, $F(1, 22) = 5.3$, $p < 0.05$, ($M = 634$ ms vs. $M = 704$ ms, respectively). This effect was further specified by a significant two-way interaction between experiment and difficulty, $F(1, 22) = 7.1$, $p < 0.05$, suggesting that RT increases from easy to difficult stimuli were more pronounced in the expression task (*Mdiff* = 49 ms vs. *Mdiff* = 30 ms).

The separate analysis for Experiment 11 (identity task) revealed no condition effect, $F(1, 11) < 1$. The interaction between condition and difficulty was not significant, $F(1, 11) = 2.9$, $p = 0.12$ (*Mdiff* = 6 ms vs. *Mdiff* = -7 ms for differences between the control and orthogonal condition for easy and difficult identities, respectively).

Overall, familiar faces were classified faster than unfamiliar ones, as suggested by the familiarity main effect, $F(1, 11) = 11.7$, $p < 0.01$ (*Mdiff* = 30 ms). This effect was further qualified by a significant two-way interaction between familiarity and expression, $F(1,11) = 10.2$, $p < 0.01$. Inspection of Figure 26 suggests that familiar faces were recognized faster when displaying a happy expression ($M = 614$ ms vs. $M = 628$ ms, for happy and angry faces, respectively). In contrast, differences for unfamiliar faces were smaller and went in the other direction ($M = 654$ ms vs. $M = 648$ ms for happy and angry faces, respectively). The expected main effect of difficulty, $F(1, 11) = 58.2$, $p < 0.001$

reflected longer RTs for less salient morph stimuli (*Mdiff* = 30 ms). There were no other significant effects.

The separate analysis of Experiment 12 (expression task) did not yield a condition effect, $F(1, 11) < 1$, (*Mdiff* = 14 ms). The critical two-way interaction between condition and difficulty did not reach significance, $F(1, 11) = 2.0$, $p = 0.19$, although there was a numerical difference pointing into the direction of a larger condition effect for difficult stimuli (*Mdiff* = 7 ms vs. *Mdiff* = 20 ms for easy and difficult expressions, respectively).

Overall, happy expressions were classified faster than angry expressions, $F(1, 11) = 9.5$, $p < 0.05$ (*Mdiff* = 36 ms). There was also a highly significant familiarity effect, $F(1, 11) = 24.8$, $p < 0.001$, reflecting faster expression classifications for familiar faces (*Mdiff* = 54 ms). The expected effect of difficulty $F(1, 11) = 59.8$, $p < 0.001$ demonstrated faster responses to easy compared to difficult stimuli (*Mdiff* = 49 ms). There were no other significant effects.

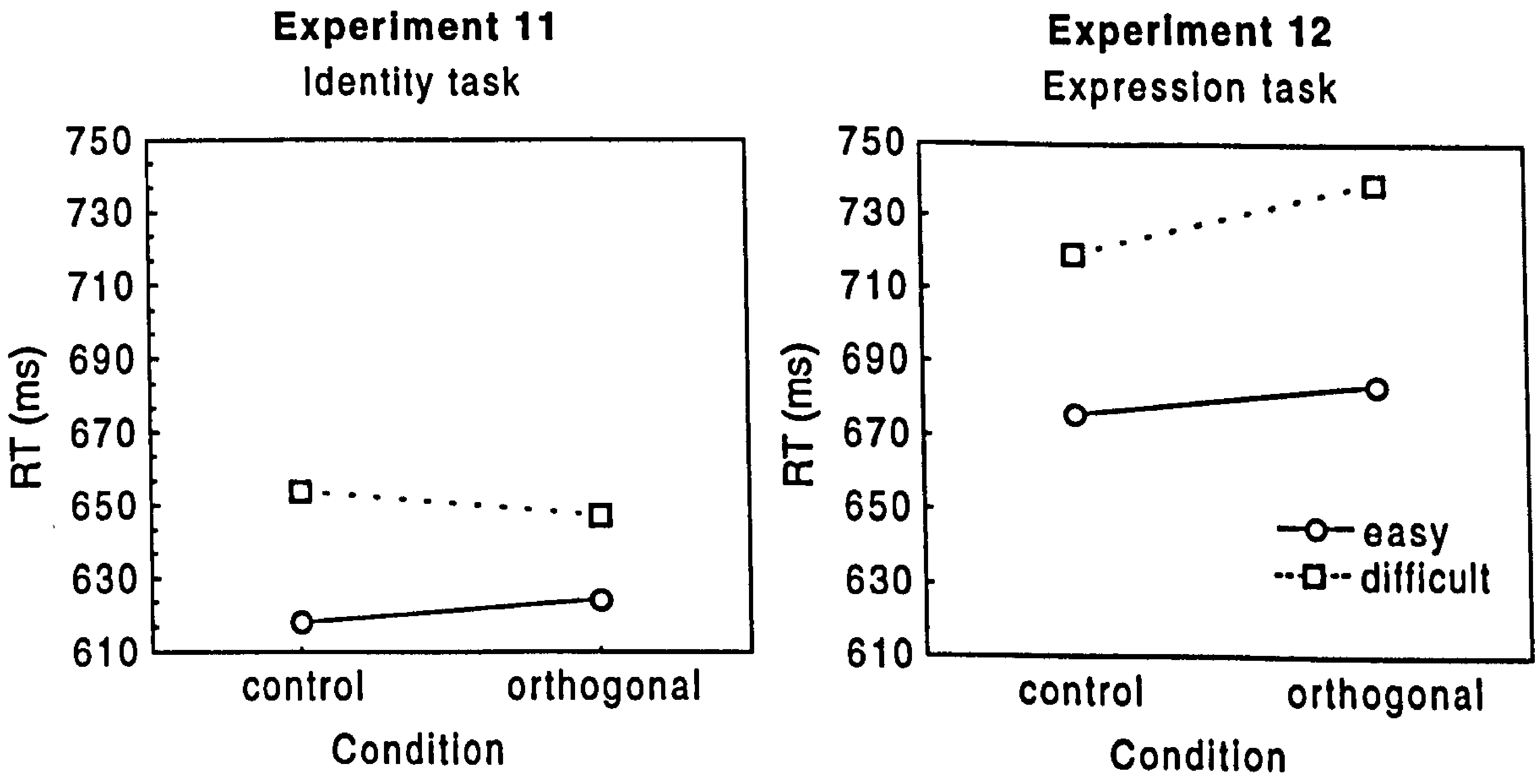

Figure 25: Reaction times in Experiments 11 and 12. Overall, there was no condition effect and no interaction between condition x difficulty of the relevant dimension. However, the interaction between experiment, condition and difficulty was significant, $F(1, 22) = 4.7, p < 0.05$.

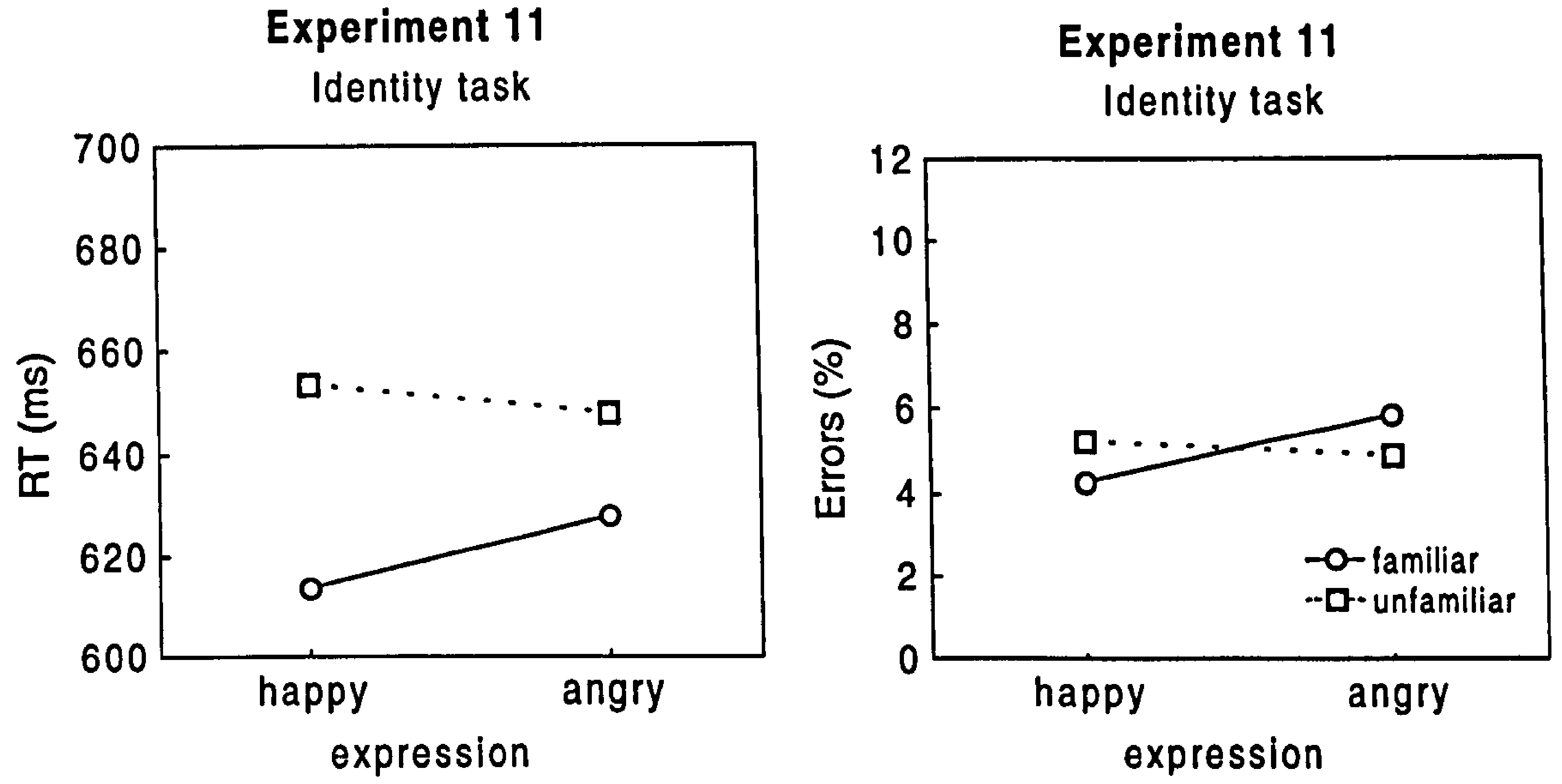

Figure 26: Two-way interaction between familiarity and displayed expression in the identity task. Familiar faces were recognized faster when they showed a happy expression. The trend for the error rates shows a similar pattern.

6.4.2 Error rates

The ANOVA across experiments showed no condition effect, F(1, 22) < 1, no interaction between condition and group, $F(1, 22) < 1$ and no interaction between condition and difficulty, F(1, 22) < 1. There was a trend for higher error rates in the expression task, $F(1, 22) = 3.2$, $p < 0.09$ and a main effect of difficulty, $F(1, 22) = 63.4$, $p < 0.001$ (see also Figure 27).

The analysis for Experiment 11 (identity task), revealed no condition effect, F(1,11) < 1, and no interaction between condition and difficulty, F(1, 11) < 1. In addition to the expected effect of difficulty, $F(1, 11) = 27.6$, $p < 0.001$, reflecting higher error rates for difficult stimuli (*Mdiff* = 4%), there was a three-way interaction between condition, familiarity and difficulty, $F(1, 11) = 4.9$, $p < 0.05$ and a four-way interaction between condition, familiarity, expression and difficulty, $F(1, 11) = 8.1$, $p < 0.05$. They were further explored by separate post-hoc ANOVAs for happy and angry faces.

Only the ANOVA for happy faces showed a three-way interaction between condition, familiarity and saliency, *F*(1, 11) = 11.4, $p < 0.01$. Inspection of Figure 28 suggests that this was due to a stronger difficulty effect for familiar faces in the orthogonal and for unfamiliar faces in the control condition.

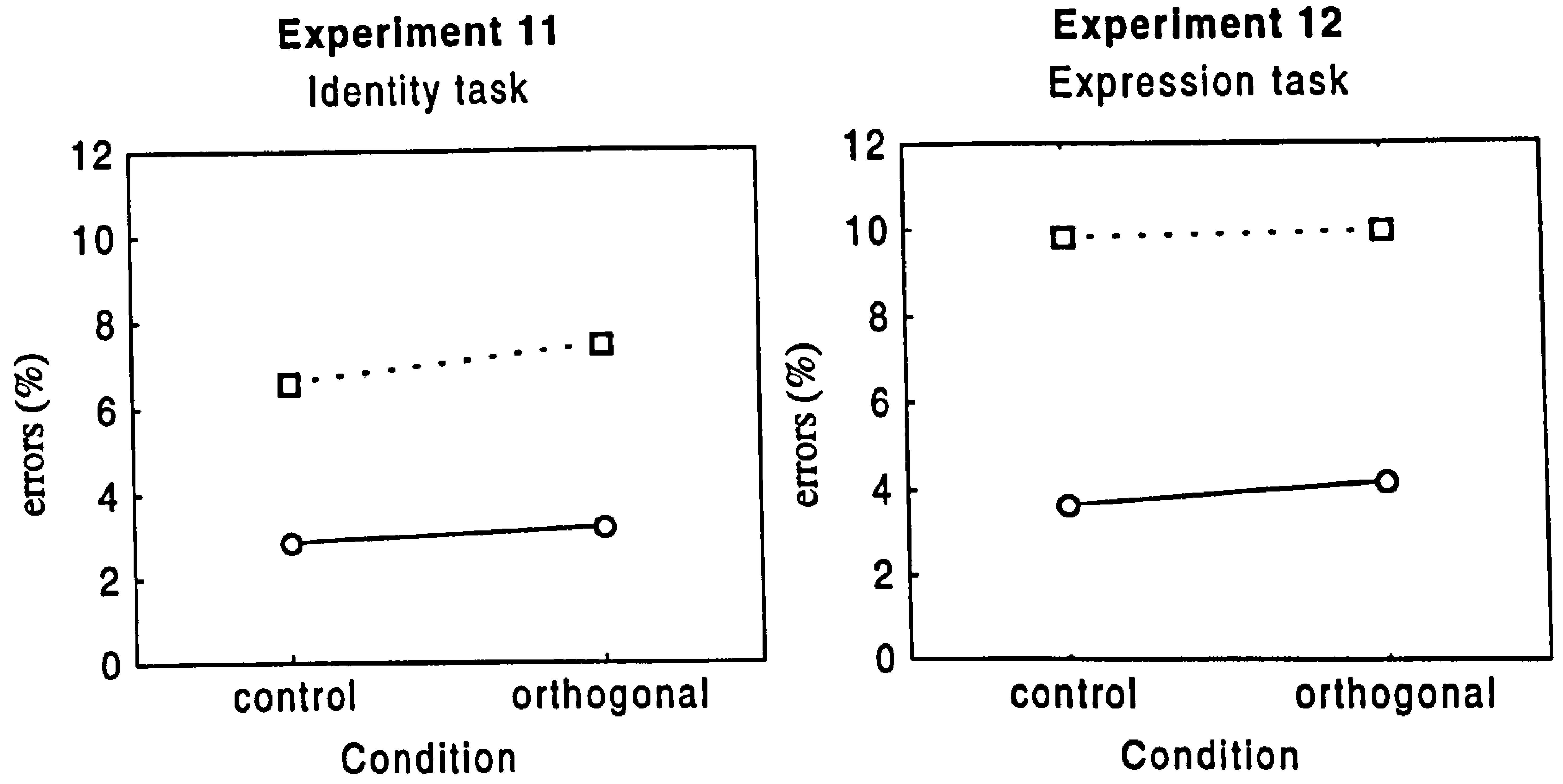

Figure 27: Error rates in Experiments 11 and 12. Overall, there was no effect of condition and no interaction between condition x stimulus saliency.

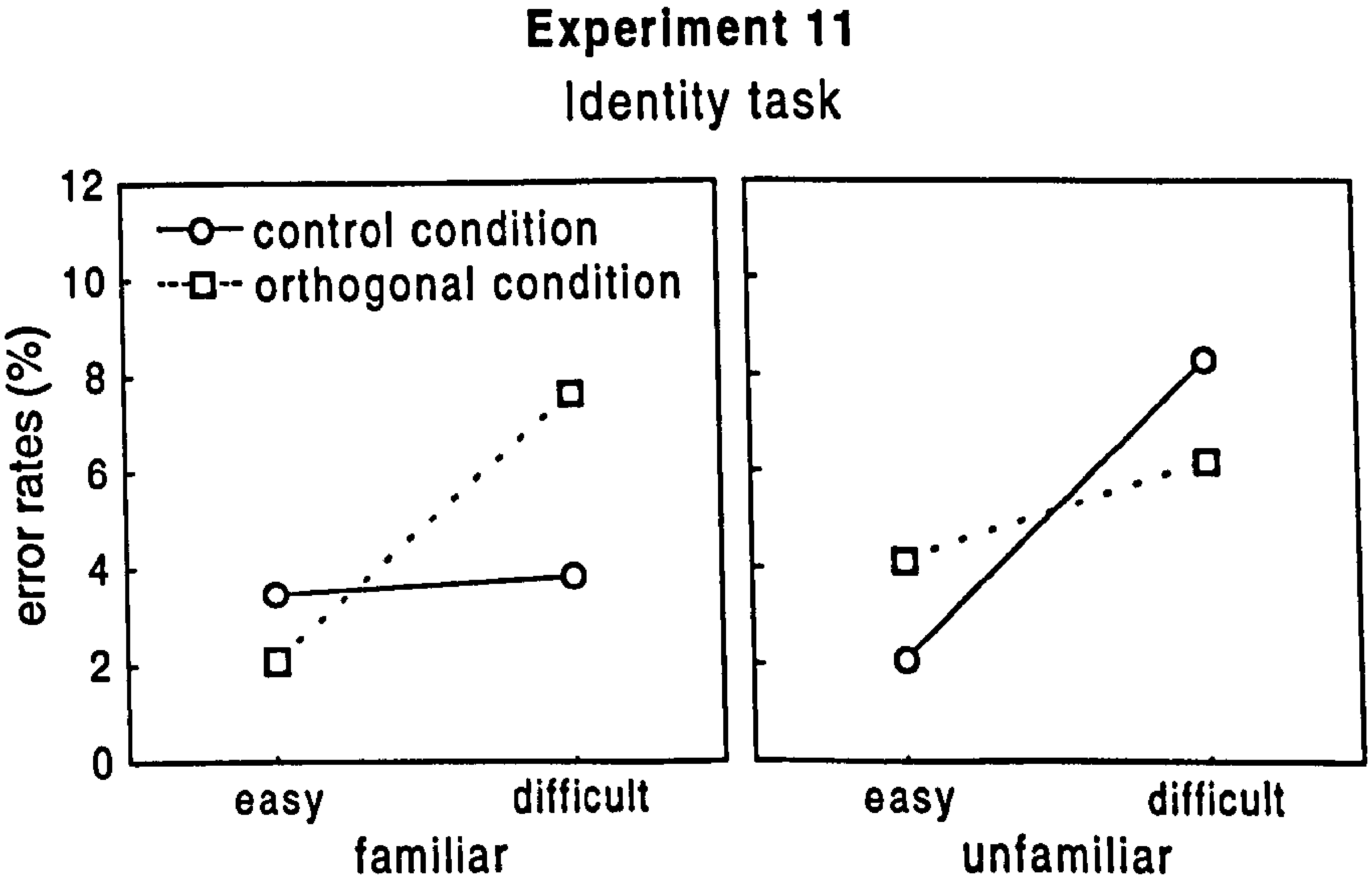

Figure 28: Error rates for Experiment 11 (identity task): three way interaction between condition, familiarity and stimulus saliency for happy faces.

For Experiment 12 (expression task), the analysis revealed no significant condition effect, F(1, 11) < 1 and no interaction between condition and difficulty, F(1, 11) < 1.

There was a significant effect of expression, $F(1, 11) = 5.9$, $p < 0.05$, reflecting higher error rates for angry faces (*Mdiff* = 2.6%). The familiarity effect $F(1, 11) = 6.2$, $p < 0.05$ showed overall higher error rates for unfamiliar faces (*Mdiff* = 2.4%). These main effects were further qualified by a two-way interaction between expression and familiarity $F(1, 11) = 30.7$, $p < 0.001$. Inspection of Figure 29 suggests that differences between familiar and unfamiliar faces were more pronounced for happy expressions (*Mdiff* = 5.3%) while error rates were very similar for angry faces (*Mdiff* = -0.5%).

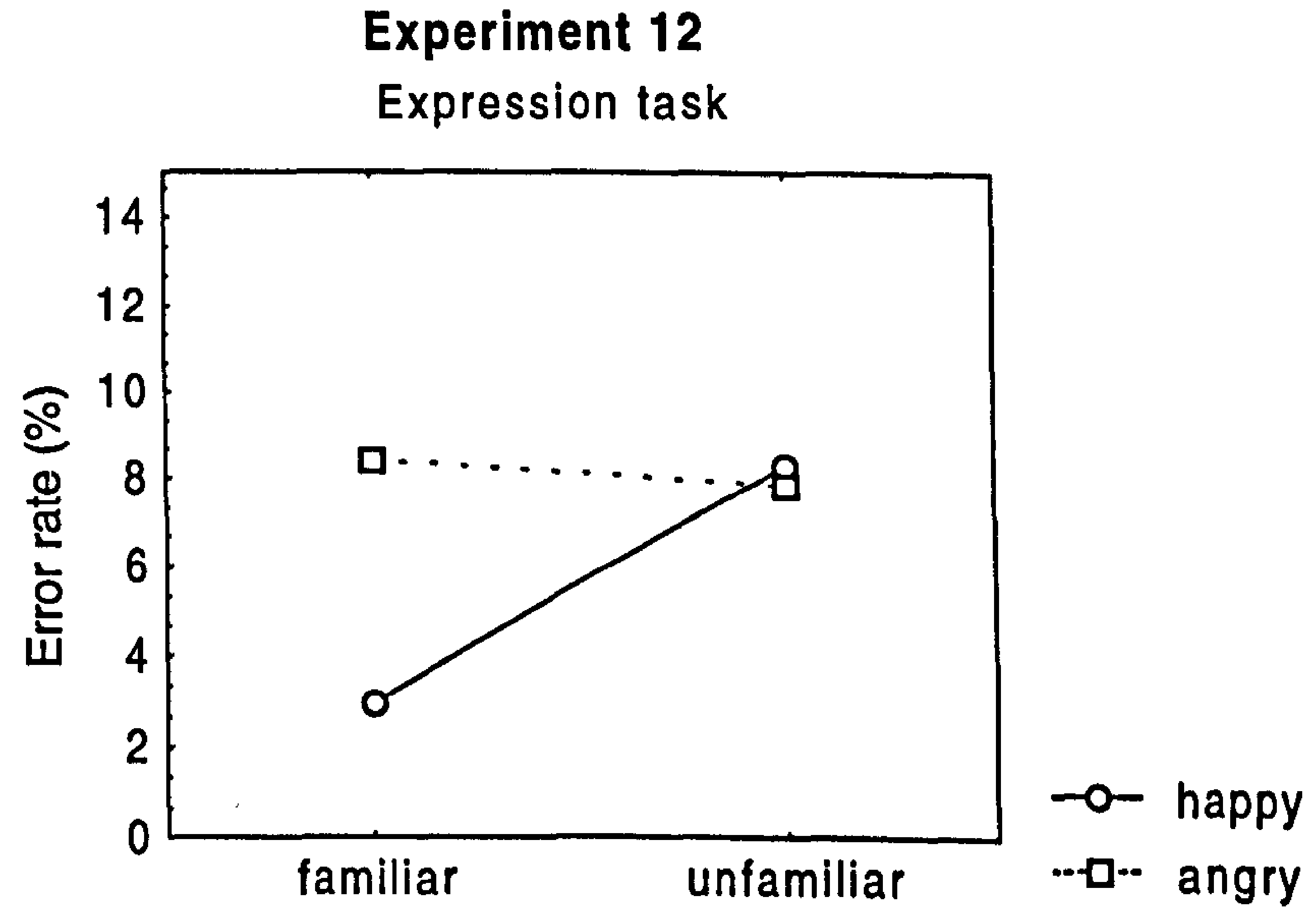

Figure 29: Error rates in Experiment 12: Differences between familiar and unfamiliar faces interacted with type of expression.

The expected effect of difficulty, $F(1, 11) = 36.1$, $p < 0.001$, reflected higher error rates for the difficult compared to the easy stimuli (*Mdiff* = 6%). This effect was further specified by a significant interaction between familiarity and difficulty, $F(1, 11) = 11.3$, p

< 0.01. Inspection of Figure 30 suggests that the difference between easy and difficult stimuli was larger for unfamiliar faces (*Mdiff* = 8.9% vs. *Mdiff* = 3% for unfamiliar and familiar faces, respectively). No other effects were significant.

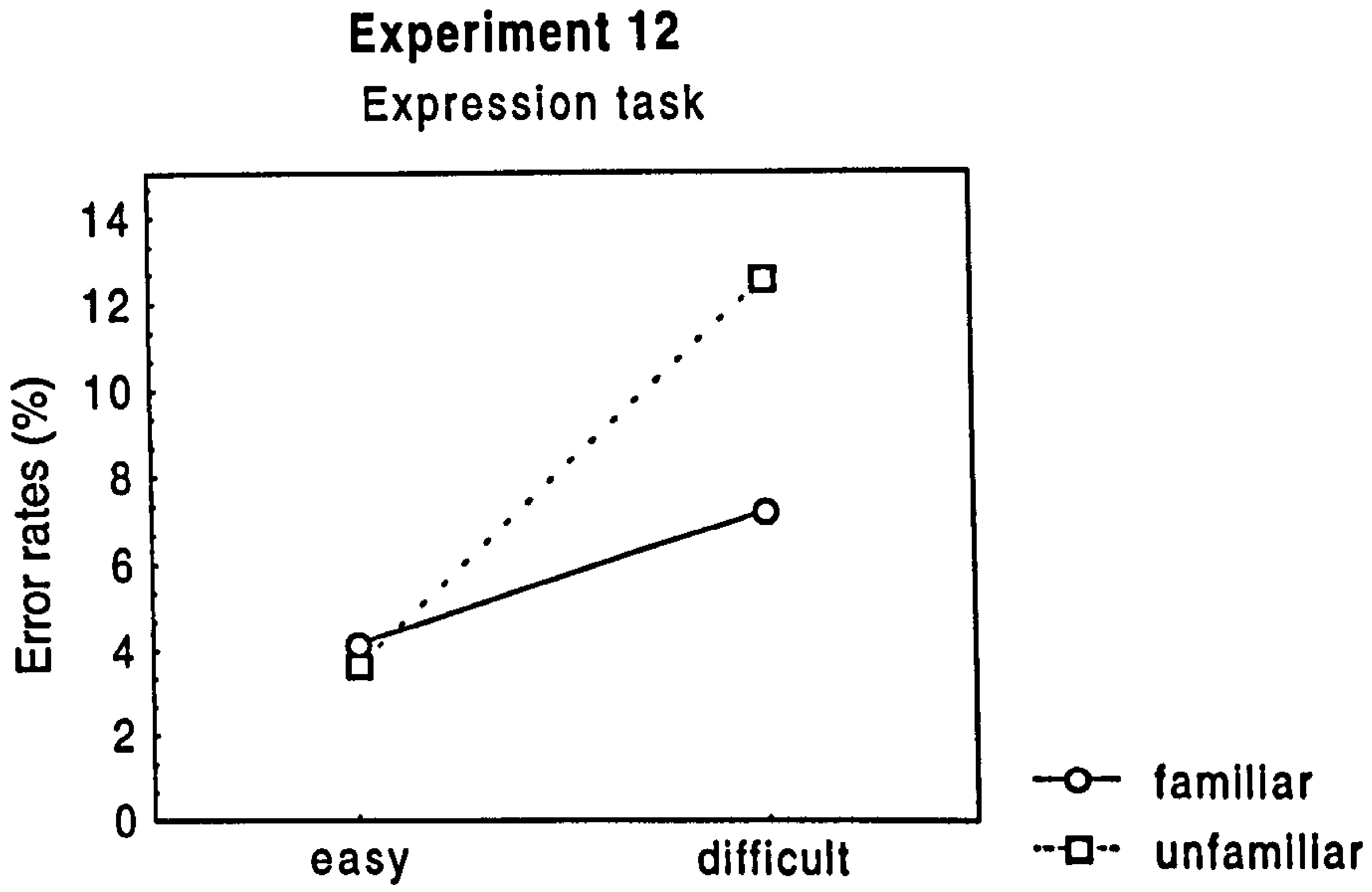

Figure 30: Error rates in Experiment 12: interaction between expression difficulty and familiarity.

6.5 Discussion

Experiments 11 and 12 further investigated the conflicting findings of an asymmetric interaction between the processing of identity and expression (Schweinberger et al., 1998) and the absence of such an interaction in Experiments 6 and 7. It was argued that the results of Schweinberger et al. (1998) might have been produced by overall differences in stimulus saliency and relative processing speed between the two dimensions. In the cited experiments, identity classifications tended to be faster than classifications of expression, while no such tendency was present in Experiments 6 and 7. This possibility was tested by selectively manipulating the perceptual saliency of the respective relevant dimension by means of morphing. In Experiment 11, identity was either easy or difficult to classify, while task-irrelevant expression information was always highly salient. Vice versa, in Experiment 12, expression was easy or difficult to classify, while task-irrelevant face identity information was highly salient. It was reasoned that an interaction between condition and difficulty in the selective attention paradigm would argue for a crucial role of relative processing speed for the reported asymmetric interaction between identity and expression processing.

As in Experiments 6 and 7, no asymmetric interaction between the identity and the expression task was found. In neither of the two experiments did task performance differ between the control and the orthogonal condition. Also, none of the separate analyses per experiment yielded an interaction between condition and difficulty. Therefore, the experiments suggest that relative processing speed was not the crucial factor for the difference between the study of Schweinberger et al. (1998) and Experiments 6 and 7.

Overall, expression classifications were slower than familiarity decisions, but no orthogonal interference was found in the expression task which means that differences in task difficulty alone do not seem to cause an asymmetric interaction between identity and expression processing, at least when identity is defined in terms of a super-ordinate familiarity category. However, there is some evidence that differences in perceptual saliency might have partly influenced orthogonal interference in the Garner task, as suggested by the significant three-way interaction between experiment, condition and difficulty for the reaction times. In the expression task, there seemed to be some interference in the orthogonal condition only for the difficult expressions, while in the identity task the difference between the control and the orthogonal condition for the difficult identities was marginal. Because even the difficult face identities were classified faster than the easy expressions, such a pattern would rather support the idea of an impact of relative processing speed in the selective attention paradigm. However, it has to be noted that the separate analysis of the expression task did not show a significant interaction between condition and difficulty, although the numerical difference between the control and the orthogonal condition was nearly three times higher for the difficult stimuli compared to the highly salient expressions. This and the significant three-way interaction between experiment, condition and difficulty might suggest that the non-significant interaction between condition and difficulty in the expression task could have been due to a lack of statistical power.

The question of an influence of perceptual saliency and relative processing speed in the selective attention paradigm has already been addressed by Schweinberger et al. (1999). The authors found that the *presence* of an asymmetric interaction between the processing of identity and expression was independent of differences in perceptual saliency. The present experiments demonstrated that the *absence* of such an interaction was independent of relative processing speed of both dimensions. This suggests that the critical factors for the differing results of Schweinberger et al. (1998; 1999) and this study most likely have to be located elsewhere.

As already outlined above, the design of the studies also differed with respect to other major aspects. In the experiments by Schweinberger et al. (1998; 1999) very small stimulus sets were presented and participants might have used picture based response strategies. In the identity task, classifications might have been completely based on external facial features (Ellis et al., 1979) or even superficial pictorial cues if these were correlated with identity. For these features, there was no increase of variability from the control to the orthogonal condition. In contrast, for classifications of expression, participants were probably forced to attend to the facial configuration to complete the task successfully. The number of facial configurations was increased from the control to the orthogonal condition in the expression task, which might have created an asymmetric increase in task difficulty (see also section 4.1). In the stimulus sets used in the present experiments, there was significantly more overall stimulus variation, because a larger number of individuals was presented and pictures of the same identities were taken from different sources. Therefore, familiarity decisions could not be made solely on pictorial cues or just by keeping two different external features in mind as might have been the case in the study by Schweinberger et al. (1998; 1999). In addition, the number of stimulus repetitions was low, which possibly reduced visual memory effects.

The second important difference between the studies refers to the definition of "identity". In contrast to the studies by Schweinberger et al. (1998; 1999), where identity was defined in terms of individual face identity, the present experiments allowed for individual face variations within familiarity categories. Although, as in the study by Schweinberger et al. (1999), overall irrelevant individual identity variation was increased in the orthogonal condition of the expression task, it is possible that within category face variations interfered with expression classifications in the control condition. Despite the fact that the number of individuals was doubled in the orthogonal condition, orthogonal interference does not necessarily have to increase linearly to overall stimulus variability (Mullenix et al., 1990).

At present, there is not yet enough empirical evidence to clearly differentiate between these two possible explanations for conflicting findings between the present results and previous reports (Schweinberger et al., 1998; Schweinberger et al., 1999). This question will be further addressed in Experiments 13 and 14.

Despite the non-significant condition effects in both tasks, the experiments provided some evidence that stored representations of face identities might be less abstract than usually thought (Bruce et al., 1986) and preserve some information about typical expressions. This finding is in accordance with Experiments 7 and 8. Faster recognition of familiar faces for happy compared to angry expressions might be due to perceptual learning (for a detailed discussion see 5.7). The identity task also revealed faster classifications of familiar faces compared to unfamiliar faces. This effect is well established (Bruce, 1986; Young et al., 1985; Young et al., 1986) and has been associated with highly automatic and fast activation of FRUs and PINs that are only availably for familiar faces. For the error rates in the identity task, there was a complex four-way interaction, which seems to be difficult to interpret.

As in Experiment 9, expressions of familiar faces were classified faster than expressions displayed by unfamiliar faces, which might argue for dependencies between expression and identity processes. However, this has to be interpreted cautiously, as the results of Experiment 10 suggested that the effect might at least be partially due to a somewhat higher overall expressiveness of the familiar picture set.

To summarize, the results suggest that inconsistencies between Experiments 6 and 7 and previous reports, which argued for a contingency of expression analysis on face identity processes (Schweinberger et al., 1998) were not mainly due to overall differences in task difficulty. A selective attention paradigm (Garner, 1974; 1976) was applied and neither in an identity, nor in a expression task orthogonal interference of the respective task- irrelevant dimension was found. Most importantly, this did not interact with relative processing speed. It was argued that the divergence between these experiments and previous reports (Schweinberger 1998; 1999) might therefore either be due to different experimental definitions of face identity or, alternatively to an efficient control of asymmetric increases of task difficulty and pictorial strategies in the present experiments. While the first explanation would allow for an interpretation of the date in the sense of a parallel-contingent processing of expression and identity as proposed by Schweinberger et al. (1998; 1999), the later would suggest that the authors' previous observations of an asymmetric interaction between identity and expression in the Garner paradigm might not reflect stable functional dependencies between both dimensions and should be interpreted with caution.

7 Experiments 13 and 14: Effects of non-face specific pictorial variation, increases of task-relevant variability and the moderating effect of overall task difficulty

7.1 Purpose of Experiments 13 and 14

Experiments 13 and 14 had three major aims. First, they further explored discrepancies between reports of an asymmetric interaction between the processing of identity and expression (Schweinberger et al., 1998; Schweinberger et al., 1999) and the results of Experiments 6, 7, 11 and 12, where no such interaction was found. It has been outlined above that the conflicting results might either be due to different definitions of identity or alternatively they might be the result of a better control of pictorial strategies and asymmetric increases of task difficulty in the present studies compared to previous reports (Schweinberger et al., 1998; Schweinberger et al., 1999). Because both explanations have considerably different theoretical implications, Experiments 13 and 14 aimed at disentangling these two possibilities. Attributing the diverging results to different definitions of *identity* in the studies would allow for an interpretation of the results in the sense of a parallel-contingent processing of identity and expression, suggesting that identity can be processed independently of expression, but not vice-versa (Schweinberger et al., 1999). In this case task-irrelevant variations of individual face identity within familiarity categories might have produced orthogonal interference in the control condition of the expression task. Alternatively, an explanation of the contrasting results in the sense of a more efficient control of potentially asymmetric increases of task-relevant information and pictorial strategies in the present experiments would suggest that previous findings of an asymmetric interaction between identity and expression processing might be at least partially stimulus dependent and might not generalize.

The second aim of the experiments was to explore an alternative explanation for the diverging results which has not been considered yet. It is possible that asymmetric interactions between identity and expression in the selective attention paradigm might be a result of an overall stronger susceptibility of expression processing to non face-specific pictorial variations. In Experiments 6, 7, 11 and 12, where overall pictorial stimulus variability between blocks was similar, no asymmetric interaction was found. It might therefore be that increases in pictorial, non face-specific variation from the control to the orthogonal condition have a stronger impact on expression than identity processing. This was investigated by including an orthogonal condition which introduced task-irrelevant pictorial variations, while the respective task-irrelevant facial dimension was held constant.

Finally, the third aim was to have another close look at the role of overall task difficulty and relative processing speed in the selective attention paradigm. Although the results of Experiments 11 and 12 did not suggest that differences in perceptual saliency were the main factor for differences between previous reports of an asymmetric interaction between identity and expression processes in the Garner paradigm (Schweinberger et al., 1998), they provided some evidence that large differences in task difficulty might modulate orthogonal interference. This is a crucial point, because there are studies suggesting that the matching of facial identity is easier than the matching of expressions, if the external facial features, such as hair, head shape and ears are provided (Münte et al., 1998; Potter et al., 1997). Importantly, there was also a trend suggesting faster classifications of identity than expression in Experiments 3 and 4 by Schweinberger at al. (1998) and Experiments 1A and 1B by Schweinberger at al. (1999).

Although no orthogonal interference of identity on expression was found in Experiment 11, where expression was overall more difficult to classify, there was some evidence for a potential influence of relative processing speed in form of a three-way interaction between task, condition and difficulty. Schweinberger et al. (1999) tried to explore the influence of relative processing speed by selectively manipulating the difficulty of the identity task by means of morphing between two identities. They claimed that expression classifications were modulated by identity variations, even when identity was more difficult to classify than expressions. However, a close inspection of Figures 2 and 3 in their study suggests that this was not the case. For the difficult identity morphs, classifications of expression still seemed to be slower than classifications of identity. Due to the categorical processing of identity (Beale et al., 1995) it is questionable whether it is possible to create morph stimuli that are classified consistently but significantly slower with respect to identity compared to expression. The same problem occurred in the present Experiments 11 and 12, where identity classifications for difficult identity morphs where still made faster than expression classifications for highly salient expressions. It was therefore necessary to find another way of making identity classifications relatively more difficult and to investigate the influences of identity variations on the processing of expressions under such circumstances.

7.2 Rationale of Experiments 13 and 14

The possibility that asymmetric interactions between the processing of facial identity and facial expression (Schweinberger et al., 1998; Schweinberger et al., 1999) are a result of a higher susceptibility of the expression dimension to non face-specific pictorial variations, was tested by including an "orthogonal pictorial" condition. In this condition, task-irrelevant non face-specific pictorial information was varied, while the respective task-irrelevant facial dimension was held constant. This was achieved by presenting the same face stimuli either in colour or greyscale mode in a "orthogonal pictorial" condition. If expression was more vulnerable to such superficial pictorial variations, performance should decrease in the "orthogonal pictorial" condition of the expression task, but remain unaffected in the identity task.

To further investigate the role of relative processing speed, stimuli were prepared that could be classified easier with respect to expression compared to identity. Experiments 8 and 9 had demonstrated that morphing can be used to selectively manipulate difficulty on the identity dimension. However, due to the categorical processing of identity, it might not be possible to produce morph stimuli that are classified consistently as belonging to a particular identity, but at the same time with significantly longer RTs compared to expression classifications. Therefore another approach was chosen in order to make the expression task easy and the identity task more difficult. First, pictures displaying highly salient expressions were obtained by hiring professional actors who posed for recordings of unambiguous emotional expressions. To make identity decisions more difficult, two strategies were followed. First, actors of similar age and general appearance were selected. Second, the faces were presented without most of the external features. It is known that especially for unfamiliar faces, external facial features play a major role for identity discrimination (Ellis et al., 1979; Young et al., 1985). In particular hair style is a very strong cue for the matching of unfamiliar faces while for familiar faces the eye region gains importance (O'Donnell et al., 2001). Removing the external features from unfamiliar faces is therefore an efficient way to selectively increase task difficulty for the identity dimension, leaving expression unaffected (Bobes et al., 2000). In the present experiments, external features such as hairstyle, overall head shape and ears were not visible in the stimuli. Most importantly, by reducing most of the external features from the stimuli, participants were forced to base both expression and identity decisions on the same facial areas. This ruled out that potential differences between condition effects in both tasks might be due to asymmetric increases in relevant information from the control to the orthogonal condition.

The fact that participants highly rely on external facial features for classifications of unfamiliar faces (Ellis et al., 1979; O'Donnell et al., 2001; Young et al., 1985) might have influenced previous results of previous studies which had applied selective attention paradigm (Schweinberger et al., 1998; Schweinberger et al., 1999). If pictures of only two individuals are presented, subjects might only concentrate on the external features in the identity task. This might make it relatively easy to ignore changes of expressions, which occur mainly in the mouth and eye region. The external features, on which identity classifications can be successfully based are not affected at all by these variations. Most importantly, this also has different consequences for the increase of *task-relevant* variation from the control to the orthogonal condition in the expression and the identity task. In the expression task, doubling the number of individuals in the orthogonal condition in order to increase *task-irrelevant* variation, also doubles the number of different expression *exemplars*, even if the number of *basic expressions* remains the same. Obviously, this also increases the amount of *task-relevant* information in the expression task. In contrast, the number of exemplars of external features that can be used for "identity" classifications is identical in the control and the orthogonal condition of the identity task. Removing the external features in Experiments 13 and 14 was used to investigate whether this difference in *task-relevant* variation might have lead to asymmetric interactions in previous reports (Schweinberger et al., 1998; Schweinberger et al., 1999).

Importantly, as in the studies by Schweinberger et al. (1998; 1999) the design encompasses a doubling of the stimulus sets from the control to the orthogonal condition. In contrast to the cited studies, the increase of task-relevant information associated with increasing stimulus set size is the same in both tasks. In addition, the design allows for an investigation of the influence of non face specific pictorial variations.

In order to increase statistical power, a repeated measurements design was used in which type of task was defined as a within subjects factor. In addition to increasing the power for detecting differences in relative processing speed between both tasks more reliably, the design allows for analysing potential effects of task order on orthogonal interference in the selective attention paradigm. This is of interest because it further addresses the question whether condition effects that are based on small sets of complex stimuli such as faces, reflect a stable functional relationship between two dimensions or might rather be influenced by an *increase* of relevant rather than an *interference* of irrelevant information. The reasoning is the following: if expression analysis is contingent on identity processing, as suggested by Schweinberger et al. (1998; 1999), and orthogonal interference caused by irrelevant identity variations in the expression task reflects the inability to selectively attend to expression due to integrated processing of both dimensions, the condition effect should not decrease with an increasing number of stimulus repetitions. The task-irrelevant identity dimension should always interfere to a similar extent with the processing of expression. Importantly, participants who perform the identity task first, will have encountered every single stimulus about fifty times before completing the expression task on the same stimuli. A smaller condition effect for these participants compared to subjects who perform the expression task first would suggest that doubling the number of stimuli in the orthogonal condition has less impact if the stimuli have already been learned before. Such a learning effect would strongly argue for a crucial role of an increase of relevant information rather than an interference of the irrelevant dimension. In contrast, a *larger* condition effect for participants who performed the identity task first might be an indicator of an integrated processing of expression and identity, possibly reflecting additional processing costs caused by switching between interdependent functions.

The following outcomes are possible: an asymmetric interaction as reported by Schweinberger et al. (1998; 1999) is observed, even when identity classifications are more difficult than classifications of expression. This would suggest that the asymmetric interaction is independent of relative processing speed and reflects a stable architecture of face recognition processes with the analysis of expression at least partly contingent on the processing of identity. In contrast, a finding of orthogonal interference only found in the more difficult task would strongly suggest that the reported asymmetric interaction in the selective attention paradigm (Schweinberger et al., 1998; Schweinberger et al., 1999) might have been caused by differences in relative processing speed.

If the asymmetric interaction in previous studies (Schweinberger et al., 1998; Schweinberger et al., 1999) was produced by a stronger susceptibility of expression processing to overall, unspecific pictorial variation, there should be a RT increase from the control to the orthogonal pictorial condition in the expression, but not in the identity task.

Finally, the finding of orthogonal interference in the orthogonal facial conditions in both tasks would suggest that the reported asymmetric interaction (Schweinberger et al., 1998; Schweinberger et al., 1999) might have been the result of an asymmetric increase of task-relevant facial information. Such an interpretation would be underlined if the condition effect in both tasks was larger, if the particular task was completed first and stimuli had not been encountered before.

7.3 Method

7.3.1 Participants

Twenty-four participants (seventeen women and seven men) aged 19–40 years (*M* = 26, *SD* = 5.1) contributed data in Experiment 13. The same participants contributed data in Experiment 14. Data for one additional participant were replaced due to an excessive overall error rate (*M* > 25%, compared to a mean of *M* = 6% across all participants).

7.3.2 Stimuli and Apparatus

The stimuli consisted of happy and angry faces of two male individuals in full frontal view presented in front of a black background. Hair and ears were covered by a black cap leaving only the internal facial features and the face shape visible. Gaze was always directed towards the camera. Both posers were professional actors from the "Stadttheater Konstanz", Germany. They were instructed to express happy and angry emotions as they would do on stage. Video clips of these expressions were recorded using a Super-VHS camera, and by means of a video capture card (AV-Master^{TM}) directly stored on the hard disk of an IBM compatible PC. Video clips were digitally edited using commercial video editing software (Ulead Media Studio^{TM}) and one frame showing the apex of each expression was selected. For each person and each expression two different exemplars were selected, which resulted in a set of eight pictures. Stimulus size and colour was adjusted using commercial graphic software (Adobe Photoshop$^{TM)}$ and portraits were saved as indexed colour (128 colours) and greyscale bitmaps, hence the complete stimulus set consisted of sixteen pictures (greyscale stimuli are depicted in Figure 31). Picture size was 10.5 by 8.5 cm at a resolution of 30 pixels/cm. The distance to the monitor was 1 m and was controlled by a chin rest. This corresponded to a vertical visual angle of 6 degrees and a horizontal visual angle of 4.8 degrees. Pictures were presented in the centre of a 19'' monitor using ERTS^{TM} presentation software (Berisoft Corporation).

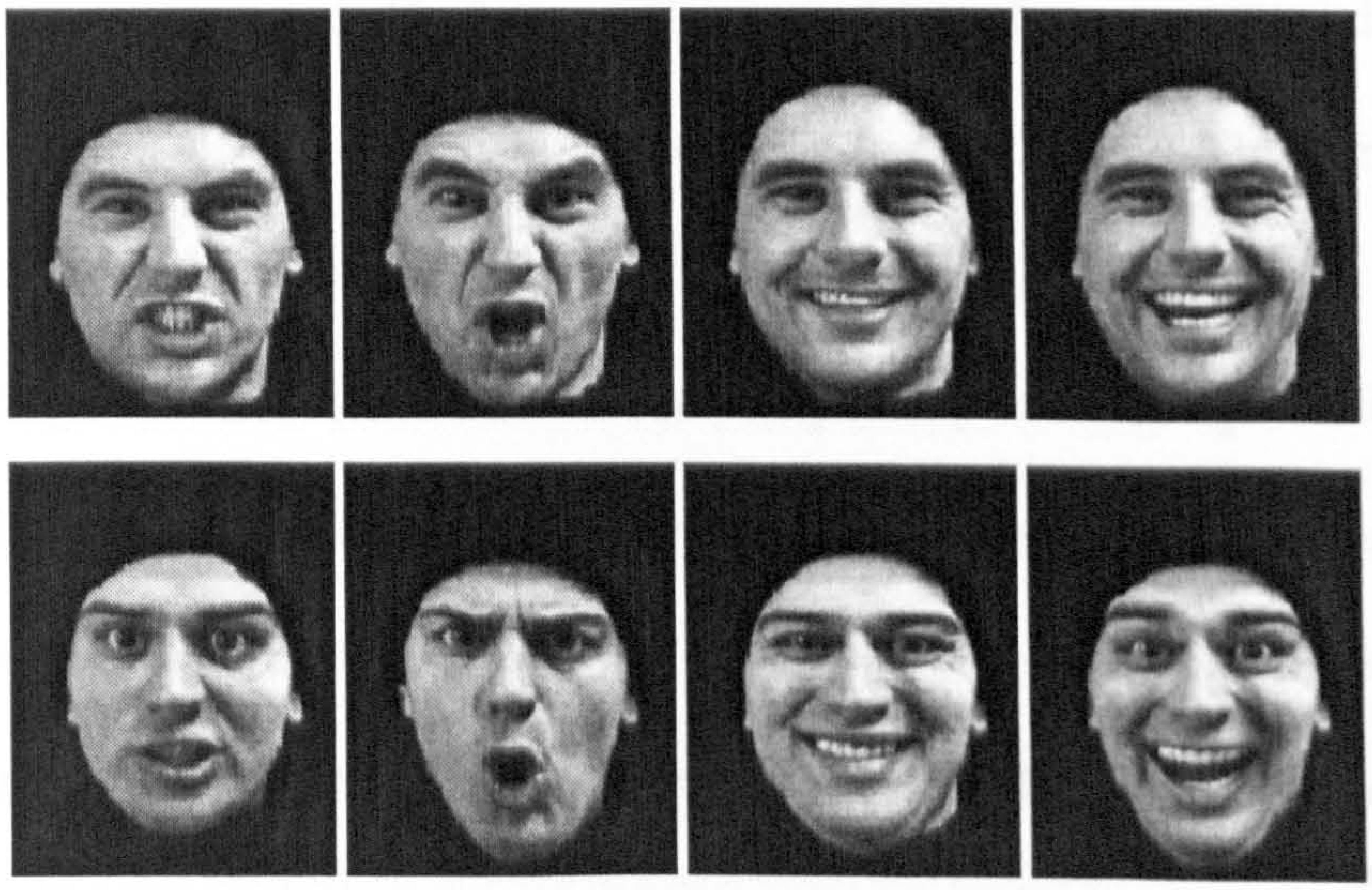

Figure 31: Stimuli presented in Experiments 13 and 14. Top row: Person A, displaying angry and happy expressions. Bottom row: Person B, displaying angry and happy expressions. The complete stimulus set consisted of these pictures plus the same images in colour.

7.3.3 Procedure

Each trial started with the presentation of a white fixation cross in the centre of the screen for 500 ms. It was replaced by a face stimulus that was visible for maximally 2000 ms or until a key was pressed. Responses with RTs below 100 ms, as well as missing or wrong answers were indicated by a tone signal presented for 140 ms. After a key press, there was a clear screen for 1000 ms. Participants responded by key presses with both hands on a special RT key pad (ERTS Key$^{TM)}$. Both speed and accuracy were stressed. The assignment of response alternative to response hand was completely counterbalanced across participants.

In Experiment 13 participants performed a Garner type speeded two choice task, and decided whether a face either showed a happy or an angry expression. In Experiment 14, participants decided whether a face either showed Person A or Person B. One half of the subjects first participated in Experiment 13 which was followed after a short break by Experiment 14. For the other half of the participants this order was reversed. Participants did not know that the relevant dimension was going to be changed. The duration of each experiment was about 16 minutes.

After reading the instructions on the monitor, participants were given four examples of the stimuli before performing eight practice trials. Both experiments consisted of three experimental conditions, labelled "control", "orthogonal pictorial" and "orthogonal facial". Conditions were further divided into blocks. Each block was preceded by eight practice trials, which were not analysed.

In the expression task (Experiment 13), one block of the control condition only showed pictures of Person A displaying both happy and angry expressions in colour. The second control block showed the same pictures of the same person but all stimuli were presented in greyscale mode. Control block three consisted of colour pictures of Person B showing happy and angry expressions. Finally, control block four presented expressions of Person B in greyscale mode. Each control block consisted of four different stimuli and every stimulus was repeated eight times per block. Each block contained 32 trials. Within blocks, stimuli were presented in random order.

In the "orthogonal pictorial" condition of Experiment 13, eight different pictures per block were presented. As in the control condition, identity was always held constant within a block, but there was additional pictorial variation caused by presenting colour and greyscale images randomly alternating within the block. Each block consisted of 64 trials and was further divided by a short break into two sets of 32 trials. Within the whole block of 64 trials, stimuli were presented in random order. On average, each of the eight stimuli per block was repeated four times per sub-block of 32 trials. The first block of the "orthogonal pictorial" condition consisted of colour and greyscale pictures of Person A, displaying both happy and angry expressions. The second block of the "orthogonal pictorial" condition showed Person B with happy and angry expressions, both in colour and greyscale mode.

In the "orthogonal-facial" condition, in addition to the relevant expression dimension, stimuli varied with respect to identity, but not colour mode. One block consisted of colour pictures of Person A and Person B displaying both happy and angry expressions, while the second block showed happy and angry expressions of both persons in greyscale mode. As in the "orthogonal pictorial" condition, each block consisted of eight stimuli and 64 trials, which were divided into sub-blocks of 32 trials. Stimuli were presented in random order within blocks and each stimulus was on the average repeated four times per sub-block.

In the identity task (Experiment 14), the control condition consisted of four different blocks in which the stimuli varied only with respect to the task-relevant identity dimension. One control block showed happy expressions of both persons in colour. The second control block showed the same expression but all stimuli were presented in greyscale. Control block three consisted of colour pictures of both persons displaying angry expressions. Finally, control block four presented angry expressions of both persons in greyscale. All blocks consisted of 32 trials and each picture was repeated eight times per block.

In the "orthogonal pictorial" condition of Experiment 14, eight different pictures per block were presented. In addition to the task-relevant identity dimension there was additional pictorial variation because colour and greyscale images were presented within the same block. As in Experiment 13, each block consisted of 64 trials and was further divided by a short break into two sets of 32 trials. Within the whole block of 64 trials, stimuli were presented in random order. On average, each of the eight stimuli per block was repeated four times per sub-block of 32 trials. The first block of the "orthogonal pictorial" condition consisted of colour and greyscale pictures of happy expressions, displayed by Person A and B. Accordingly, the second block of the "orthogonal pictorial" condition showed angry expressions of both actors, both in colour and greyscale mode.

The "orthogonal-facial" condition in Experiment 14 was identical to the one in Experiment 13. Within blocks, the stimuli varied with regard to identity and expression, but colour mode was always held constant within a block.

The order of experimental conditions was completely counterbalanced across participants. The order of sub-blocks within conditions was counterbalanced orthogonally to the order of conditions. This design made sure that in both tasks and across all conditions, the same stimuli were shown and each single stimulus was repeated eight times per condition, making sure that potential differences between conditions were not due to overall stimulus differences. However, as in the design used by Schweinberger et al. (1998; 1999), control sub-blocks contained *four* different stimuli, while sub-blocks in the orthogonal facial condition consisted of *eight* different stimuli.

7.4 Results

Misses and outliers (RT < 100 ms or > 1500 ms) were very rare in both tasks (*M* = 0.1% and *M* = 0.2% in Experiment 13; *M* = 0.2% and *M* = 0.4% in Experiment 14, respectively) and were not further analysed. Both for reaction times and errors of commission initial ANOVAs with repeated measurements on the factors experiment (expression vs. identity task), condition (control vs. orthogonal pictorial vs. orthogonal facial), person (A vs. B), expression (happy vs. angry) and mode (colour vs. greyscale) were performed. Then, separate analyses were performed per experiment, including the repeated measurement factors condition (control vs. orthogonal pictorial vs. orthogonal facial), person (A vs. B), expression (happy vs. angry) and mode (colour vs. greyscale).

7.4.1 Reaction times

The ANOVA including data from both experiments showed a significant effect of experiment, $F(1, 23) = 7.5$, $p < 0.05$, suggesting that overall, identity classifications were more difficult than expression classifications (*M* = 552 ms vs. *M* = 510 ms). The effect of experimental condition was highly significant, $F(2, 46) = 7.5$, $p < 0.01$. According to Duncan's Multiple Range post-hoc tests ($\alpha = 0.05$), RTs in the orthogonal facial condition were significantly higher than RTs in both the control and the orthogonal pictorial condition (*M* = 546 ms vs. *M* = 525 ms and *M* = 522 ms in the orthogonal facial, control and orthogonal pictorial conditions, respectively, see also Figure 32). The critical two-way interaction between experiment and condition was not significant, *F*(2, 46) = 1.3, *p* = 0.3.

The separate analysis of Experiment 13 (expression task) yielded no significant effect of condition, *F*(2, 46) = 1.4, *p* = 0.26 (Ms = 506 ms, 506 ms and 519 ms for the control, orthogonal pictorial and orthogonal facial condition, respectively). There was a significant main effect of expression, *F*(1, 23) = 11.2, *p* < 0.01, indicating shorter RTs for classifications of happy compared to angry faces (*Mdiff* = 20 ms). There was a marginally significant two-way interaction between condition and mode, *F*(2, 46) = 3.1, p = 0.06.

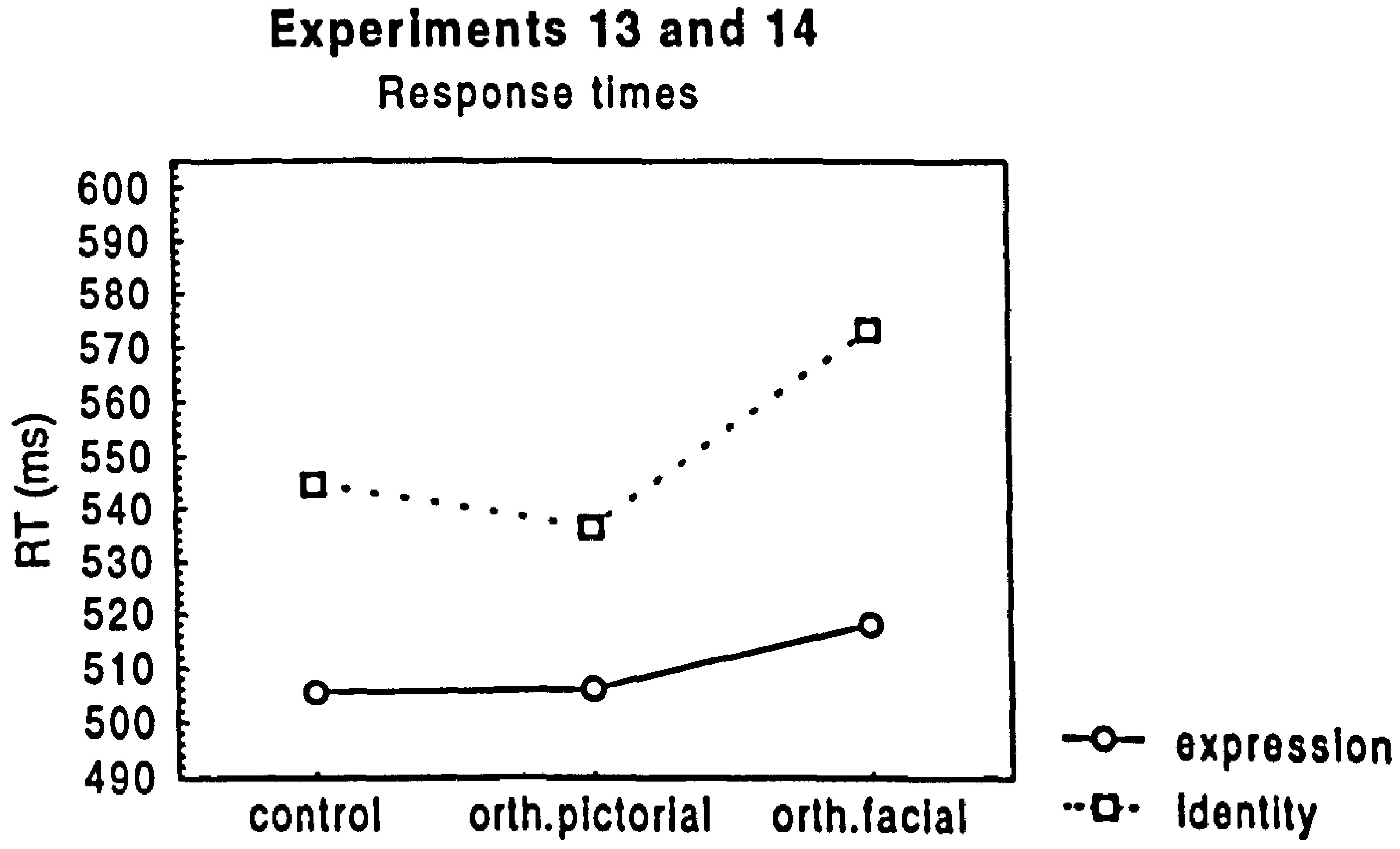

Figure 32: Mean Reaction Times in Experiment 13 (expression task) and Experiment 14 (identity task). There was no significant interaction between condition and experiment.

The separate analysis of Experiment 14 (identity task) yielded a highly significant effect of condition, *F*(2, 46) = 5.7, *p* < 0.01, which was further investigated by post-hoc tests (Duncan's multiple Range Test, $\alpha = 0.05$). According to these, RTs were reliably shorter in the control and the orthogonal pictorial condition in comparison to the orthogonal facial condition (*M* = 545 ms, *M* = 537 ms and *M* = 573 ms, for the control, orthogonal pictorial and orthogonal facial condition, respectively). The main effect of person, $F(1, 23) = 9.3$, $p < 0.01$, suggested overall faster identity classifications of Person B compared to Person A (*Mdiff* = 19 ms). This effect was further modulated by the two-way interaction between person and expression, $F(1, 23) = 9.9$, $p < 0.01$. Inspection of Figure 33 suggests that Person B was recognized faster than Person A when displaying an angry expression (*Mdiff* = 36 ms), while differences between both persons were small for happy expressions (*Mdiff* = 2 ms). The highly significant three-way interaction between condition, person and expression, $F(2, 46) = 12$, $p < 0.001$, was further explored by separate ANOVAs per condition (at a Bonferroni corrected α - level of 0.016). Both in the control condition, $F(1, 23) < 1$, and the orthogonal pictorial condition, $F(1, 23) = 5.6$, $p > 0.016$, the interaction between person and expression was not significant. However, the interaction was highly significant in the orthogonal facial condition, $F(1, 23) = 23.7$, $p < 0.001$ (see also Figure 33). There were no other significant effects.

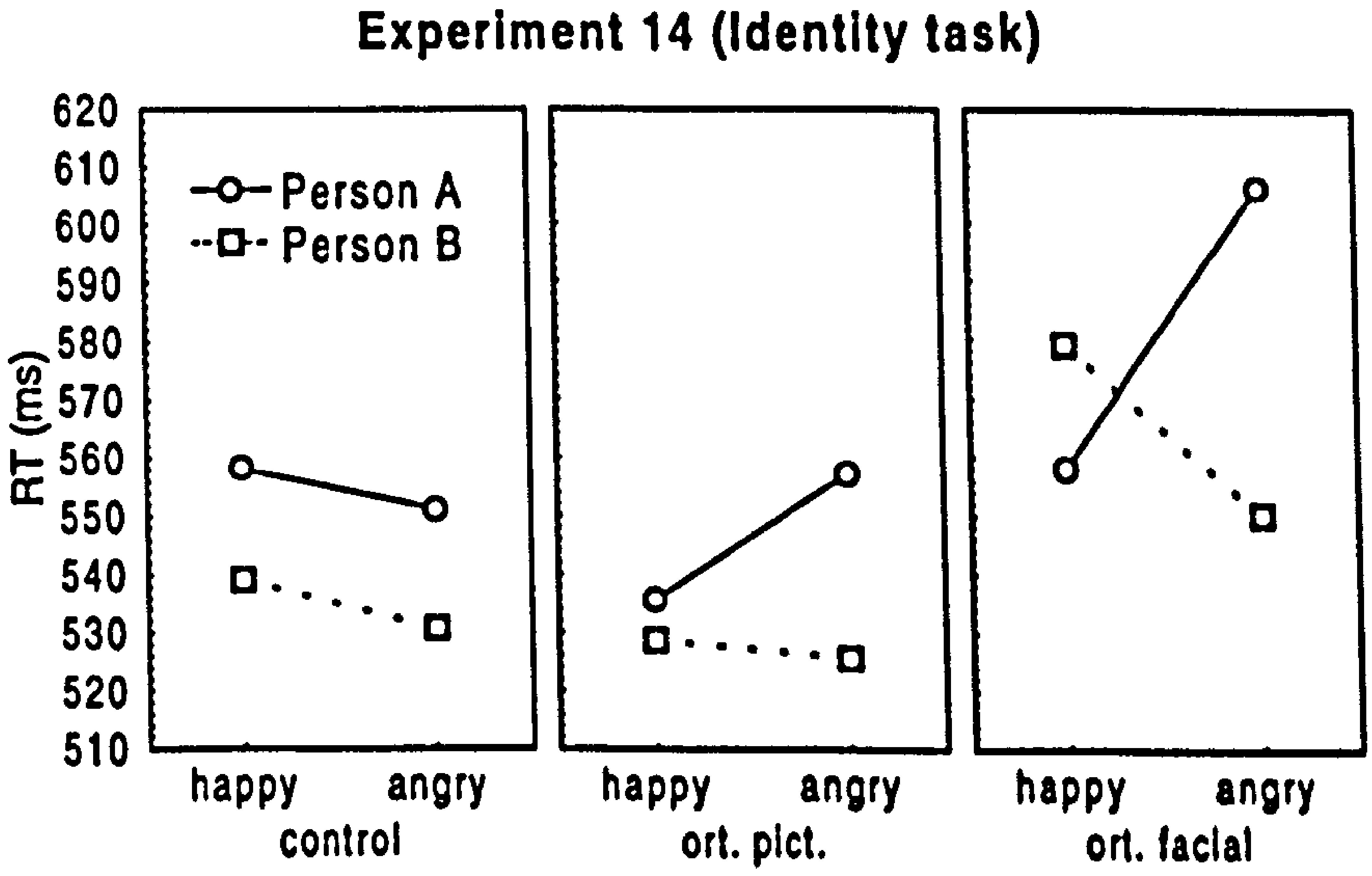

Figure 33: Three-way interaction between condition, person and expression, $F(1, 23) = 9.9$, $p < 0.01$ in Experiment 14.

7.4.2 Error rates

An ANOVA across both tasks revealed significantly higher error rates for identity classifications, $F(1, 23) = 5.1, p < 0.05$, (6.7% vs. 5.2%, respectively).

The separate ANOVA for Experiment 13 (expression task) showed no significant effects or any signs of a speed accuracy trade off.

The separate ANOVA for Experiment 14 (identity task) revealed a two-way interaction between person and expression, $F(1, 23) = 17.5, p < 0.001$, which pointed into the same direction as the RT effect and suggested that Person A was better recognized when showing a happy expression ($M = 5.1\%$ vs. $M = 8.7\%$), while the opposite was the case for Person B ($M = 8\%$ vs. $M = 5.5\%$). As for the RTs, this was further modulated by a three-way interaction between condition, person and expression, $F(2, 46) = 6.2, p < 0.01$. Post-hoc ANOVAs for each condition only yielded a two-way interaction between person and expression in the orthogonal facial, $F(1, 23) = 14.40, p < 0.001$, but not in the control, $F(1, 23) = 3.4, p = 0.08$, or the orthogonal pictorial condition, $F(1, 23) = 2.5, p = 0.13$. (see also Figure 34).

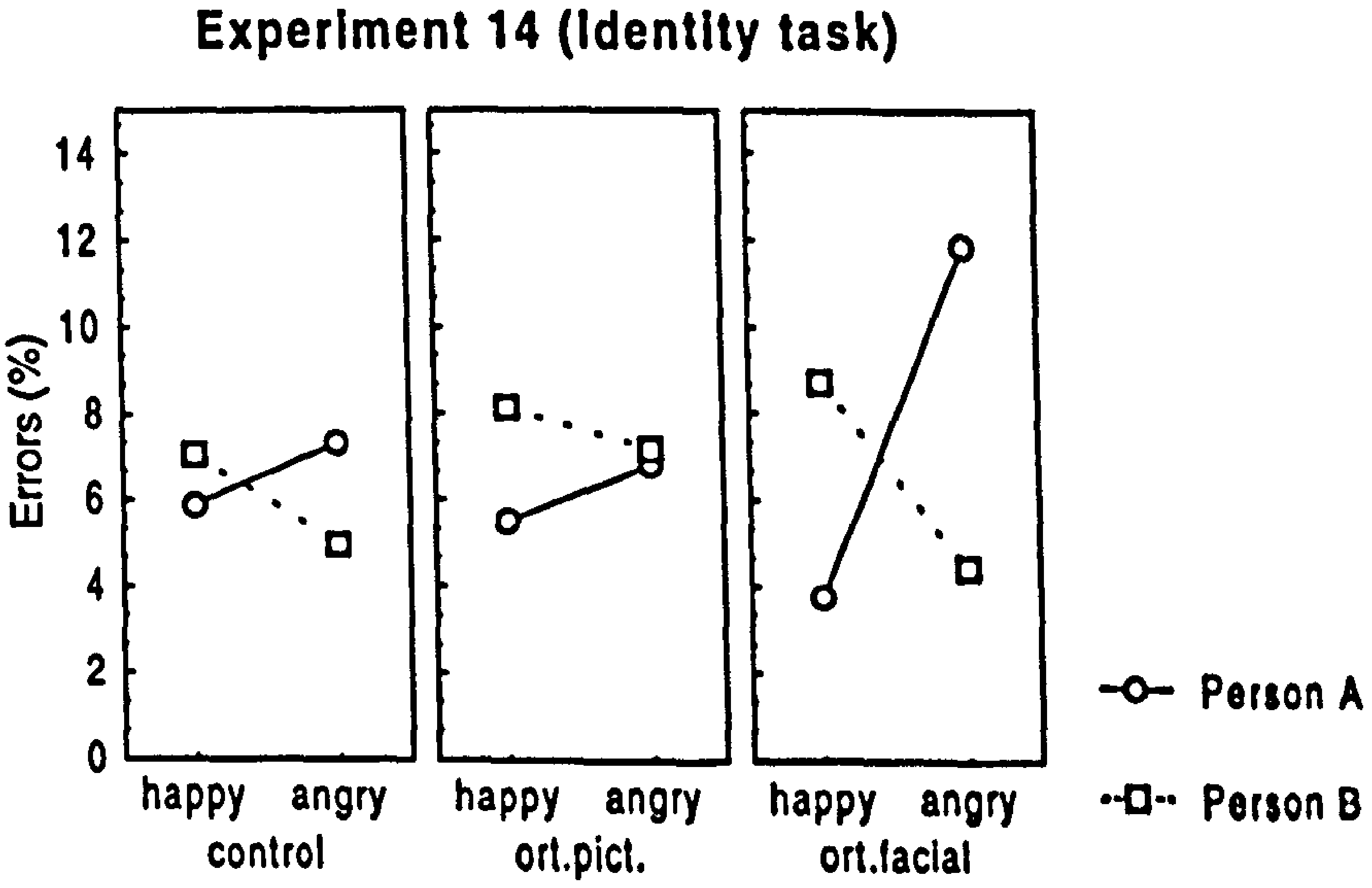

Figure 34: Error rates in Experiment 14: similarly to RTs, the two-way interaction between person and expression was only significant for the orthogonal facial condition.

7.4.2.1 Influences of task order on effects of condition

The question whether the condition effect was influenced by stimulus novelty was addressed by performing additional ANOVAs on RTs. For each experiment, data of participants who completed either the identity or the expression task first were analysed separately. Because the main interest of the analyses lies on the comparison between the control and the orthogonal facial condition and the previous analyses had shown that task-irrelevant pictorial variations did not contribute significantly to the condition effect, the orthogonal pictorial condition was not included in these analyses. The ANOVAs encompassed the repeated measurement factors condition (control vs. orthogonal facial), person (A vs. B), expression (happy vs. angry) and mode (colour vs. greyscale).

In Experiment 13 (expression task) there was a significant condition effect, $F(1, 11) = 8.2$, $p < 0.05$, for subjects who completed the expression task first (*Mdiff* = 27 ms, see also Figure 35). In contrast, for participants who had performed the identity task first, there was no indication for a condition effect, $F(1, 11) < 1$, (*Mdiff* = 0 ms).

In Experiment 14 (identity task), there was a large numerical but only marginally significant condition effect for participants, who completed the identity task first, $F(1,11) = 3.8$, $p = 0.08$ (*Mdiff* = 39 ms, see also Figure 35). For subjects, who performed the expression task first, the condition effect was not significant, $F(1, 11) = 2.4$, $p = 0.15$ (*Mdiff* = 19 ms).

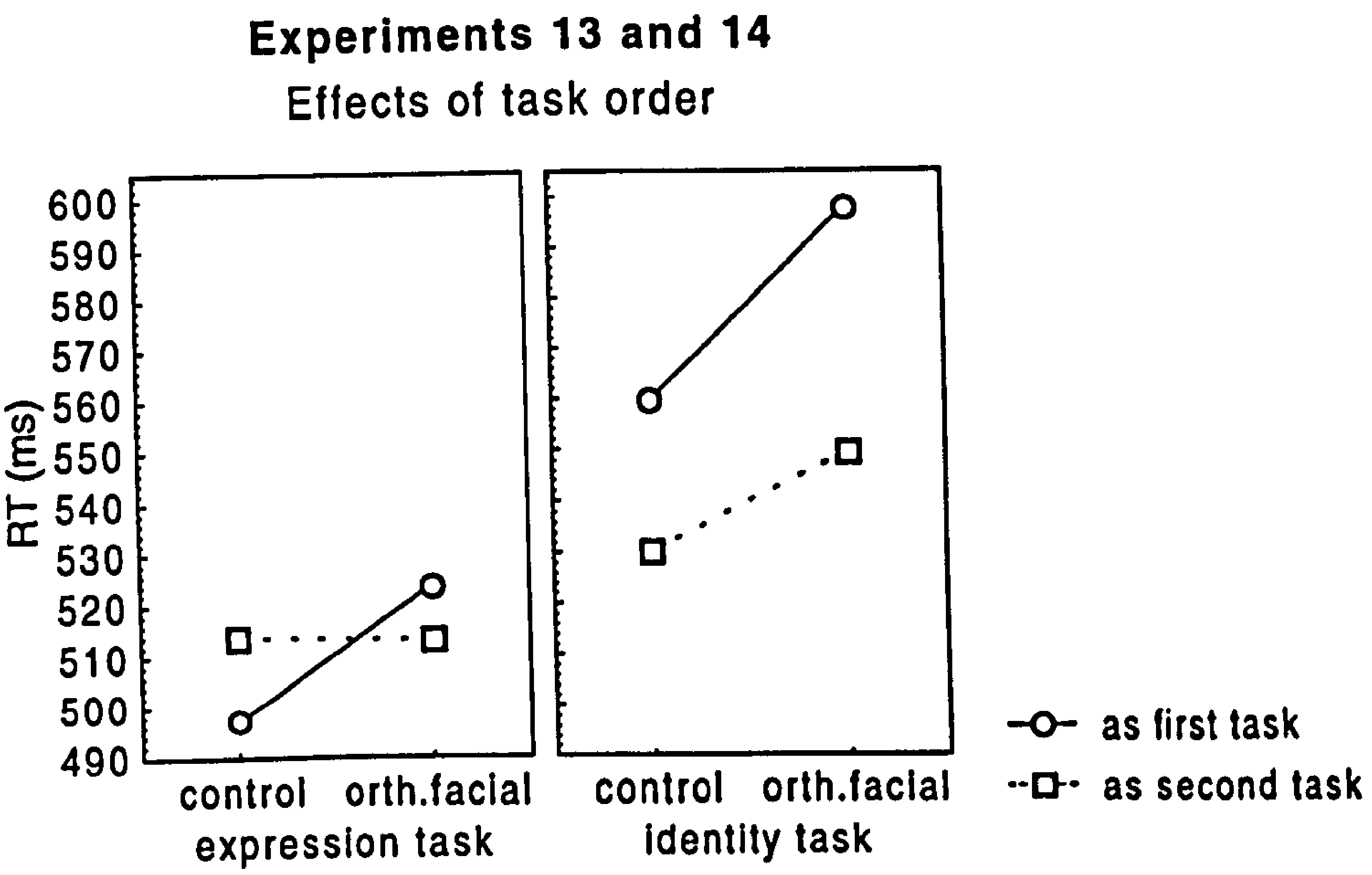

Figure 35: Comparison of RT differences between the control and orthogonal facial condition in Experiment 13 and Experiment 14, depending on task order. Note that task-irrelevant variations of previously attended dimensions could be ignored easier.

7.5 Discussion

Experiments 13 and 14 further explored discrepancies between previous reports of an asymmetric interaction between the processing of identity and expression (Schweinberger et al., 1999; Schweinberger et al., 1998) and the absence of such an interaction in Experiments 6, 7, 11 and 12.

The present experiments, which also applied the Garner paradigm of selective attention (Garner, 1974; 1976) had three major aims. The first was to investigate the possibility that the processing of expression might be more vulnerable to non face-specific pictorial variations than the processing of identity. A stronger interference of overall *pictorial variations* on expression processing might cause an asymmetric interaction in the selective attention paradigm if stimulus set size differs between the control and the orthogonal condition, as was the case in the studies by Schweinberger et al. (1998; 1999). This was explored by including a orthogonal pictorial condition, in which colour information varied while the particular task-irrelevant facial dimension was held constant.

The second aim was to have another close look at possible influences of task difficulty and relative processing speed in the selective attention paradigm. In studies that found an interference of identity variations on expression processing, identity tended to be the easier dimension (Schweinberger et al., 1998; Schweinberger et al., 1999). To my knowledge, no study so far has investigated interferences between both dimensions using stimuli that were significantly easier to classify with respect to expression. In order to make the expression task relatively easier, it was aimed at producing pictures of maximally salient facial expressions. At the same time, it was tried to selectively increase the difficulty of the identity task by removing external features and by presenting faces of two individuals of similar age and general appearance.

Finally, the third aim was to explore, whether the asymmetric interaction between identity and expression (Schweinberger et al., 1998; Schweinberger et al., 1999) might have been the result of asymmetric increases of *task-relevant facial variations.* It has been shown that in studies, which controlled for an increase of relevant facial information no orthogonal interference was found (see chapters 4 and 6). The open question was whether interference could be found both for the processing of expression *and* identity in the orthogonal condition if it was made sure that participants attend to the same facial areas in both tasks. This was investigated by applying the same paradigm and the same number of stimuli as Schweinberger et al. (1998; 1999), but presenting the faces without external features. It was reasoned that task-irrelevant variations of the mouth and eye region caused by changing facial expressions might be neglected easily in an identity task, when the external features are provided and participants adopt a strategy of only attending to the external features that are unaffected by expression variations. It was further reasoned that this might be particularly critical for very small stimulus sets, which might encourage picture based strategies especially in the identity task. It was argued that such a strategy in combination with small stimulus sets might have led to an asymmetric increase in task difficulty in previous studies (Schweinberger et al., 1998; Schweinberger et al., 1999). In contrast to the cited experiments, here the increase of *task-relevant* variation from the control to the orthogonal facial condition was comparable in both tasks, because in both experiments the number of different facial configurations, which coded the particular relevant dimension in both tasks was doubled in the orthogonal facial blocks.

The data clearly show that non-facial pictorial variations did not influence performance in either task. The comparison between the control and the orthogonal pictorial condition was not significant in the overall or separate analyses for the identity and the expression tasks. This suggests that previous reports of an asymmetric interaction between identity and expression processing (Schweinberger et al., 1998; Schweinberger et al., 1999) were not caused by a higher susceptibility of expression processing to *any* kind of non face-specific pictorial variation. Notably, a similar finding has been reported for the relationship between speechreading and identity classifications (Yakel et al., 2000).

Overall, the difference between the control and the orthogonal facial condition was highly significant. However, in contrast to previous reports (Schweinberger et al., 1998; Schweinberger et al., 1999), there was no asymmetric interaction between the processing of identity and expression. If anything, the pattern tended to be reversed. Although the interaction between task and condition was not significant, separate analyses for each experiment only showed a reliable RT increase from the control to the orthogonal facial condition in the identity, but not in the expression task when data from all participants were included.

In contrast to the abovementioned studies, classifications of identity were overall more difficult than classifications of expression, as shown by longer RTs and higher error rates. The results therefore imply that orthogonal interference in the Garner paradigm might be moderated to a considerable extent by task difficulty and might be greater for classifications of the relatively more difficult dimension. In previous studies that argued for a parallel-contingent processing of identity and expression based on findings of asymmetric interactions between both dimensions, RTs tended to be slower for expression classifications (Schweinberger et al., 1999; Schweinberger et al., 1998). The numerical differences between RTs in both tasks were not significant in those studies, but they were comparable in size to the RT differences between the expression and the identity task here. In contrast to the between-subjects design used in the cited studies, task type was defined as a repeated measurement factor in Experiments 13 and 14. It is most likely that for the stimuli of the mentioned studies, classifications of expression were more difficult than classifications of identity, but due to the reduced statistical power of the between-subjects design, the comparison of the experiments only revealed a trend.

The finding of an influence of relative processing speed on orthogonal interference in the Garner paradigm is indirectly in line with Experiments 6 and 7, where no differences with respect to task difficulty and no orthogonal increase for any of the both dimensions was found. However, the conclusion that it is relative processing speed alone, which modulates the condition effect in the selective attention paradigm is not completely supported by observations in Experiments 11 and 12, which provided only weak evidence for orthogonal interference in the more difficult expression task. Most importantly, although for the stimuli in Experiments 13 and 14 expression was easier to classify than identity, there was also a significant condition effect for those participants who performed the expression task first.

Does the overall condition effect, which did not interact with experiment and which was also present in a subgroup of participants in the expression task although identity was the “slower” dimension, imply that identity and expression are processed in an integrated manner? A detailed look at the results does not support this conclusion. The experiments rather provide evidence that an increase of task-relevant information might have caused the overall condition effect. By doubling the stimulus set, there was an increase of relevant information in both tasks. Because no external features were provided, this could possibly not be compensated in the identity task, as it might have been the case in previous studies reporting an symmetric interaction between expression and identity (Schweinberger et al., 1998; Schweinberger et al., 1999).

The analyses of task order effects also points in this direction. There was a significant condition effect in the expression task for participants who had not seen the stimuli before. Although the identity dimension was more difficult, there was a RT increase caused by doubling the number of stimuli presented in the orthogonal facial condition. Interestingly, this effect completely disappeared in subjects who had encountered the stimuli before while performing the identity task. Similarly, for classifications of identity, the numerical difference between the control and the orthogonal facial condition (which was however not significant for each group of participants taken alone) also decreased, when participants had performed the expression task first and where familiar with the stimuli. However, as opposed to the expression task, the numerical effect did not completely disappear. This pattern might suggest that the condition effects were not mainly caused by interferences of the particular task-irrelevant dimension, but by increases of task-relevant information due to doubling the stimulus sets in the orthogonal condition. Most importantly, doubling the number of stimuli seemed to have smaller effects when these had already been learned.

Task difficulty might act as moderating variable in the sense that increasing the relevant information per block might have a stronger effect on the more difficult dimension, which in this particular case was identity. This interpretation is in line with findings from Schweinberger et al. (1998; 1999) and the outcomes of the present experiments. In Experiment 6 and 7, where asymmetric increases from the control the orthogonal condition were controlled for and overall task difficulty was the same, no condition effects were found. Similar results were found in Experiments 11 and 12, although expression was overall more difficult than identity. Another additional argument against an integrated processing of both dimensions is the finding that in both experiments the respective task irrelevant condition could be even better ignored, if it had been attended to previously. This clearly argues against a stable interference effect due to contingencies between expression and identity and underlines the potential role of stimulus based effects in recent studies (Schweinberger et al., 1998; Schweinberger et al., 1999).

An alternative way to test whether orthogonal interference in the present tasks might have been due to the fact that relevant stimulus variation increased from the control to the orthogonal conditions might have the following design: a large number of expressive portraits of two individuals is presented in the selective attention paradigm. In each control block of the expression task only pictures of one individual are shown while in the orthogonal condition pictures of both face identities are presented. Keeping the number of physically different stimuli constant across conditions would control for the possibility of increasing task-relevant variability. In the identity task, the number of expression categories is doubled in the orthogonal condition, while the number of different stimuli is held constant. Ideally, only the internal facial features should be provided. A similar experiment has been carried out recently at the University of Glasgow (however, a part of the stimulus set also included external features), and orthogonal interference was neither found in the identity nor in the expression task (Bindemann, personal communication).

The finding of a significant interaction between person and expression in the orthogonal condition of the identity task might provide further evidence for expression dependent representations of faces. However, all faces were presented displaying each expression with equal frequency, so that the argument of perceptual learning is not valid. Possibly, a particular expression differentially stressed a distinctive feature of a particular face (e.g. big eyes of the happy Person A).

To summarize, the following conclusion can be drawn: processing costs in the orthogonal condition of the selective attention paradigm applied to the processing of facial identity and facial expression do not seem to be caused by unspecific pictorial variation. The effects of face related variations can be significantly modulated by relative task difficulty and especially the more difficult dimension seems to be in particular prone to influences of additional variation. The experiments have also shown that the question, which facial dimension is the more difficult one highly depends on the used stimulus material and is not a fixed characteristic.

It seems to be at least highly questionable whether the interference in the orthogonal condition of the selective attention paradigm was mainly caused by variations of the irrelevant facial dimension, because doubling the number of exemplars of such complex stimuli as faces also affects the variability of the task-relevant dimension. Previous assumptions on a parallel-contingent processing of facial expression and facial identity based on the Garner paradigm and the use of small stimulus sets (Schweinberger et al., 1998; Schweinberger et al., 1999) should therefore be interpreted with caution. The present study rather seems to imply that the cited observations might have been due to the combined influences of differences in overall task difficulty and asymmetric increases of task-relevant information from the control to the orthogonal condition.

8 General Conclusion

Fourteen experiments investigated the functional relationship between the processing of facial identity, emotional expression and facial speech, which is currently the subject of a vivid debate. Specifically the widely accepted notion of a parallel modular processing of these functions (Bruce et al., 1986), which has in particular received support from clinical neuropsychology (see e.g. Young, 1998) has been challenged recently by experimental studies with healthy participants (e.g. Walker et al., 1995; Yakel et al., 2000; Schweinberger et al., 1998; Schweinberger et al., 1999, see also section 1.2.6). A more recent model of face perception tries to integrate findings mainly based on fMRI and other brain imaging studies and suggests different neural structures as biological basis for face related processes (Haxby et al., 2000). Importantly and in contrast to Bruce and Young (1986), the authors explicitly consider the possibility of functional interactions between the supposedly involved structures. It was the aim of the present experiments to further explore this possibility.

In order to differentiate between modular and related face perception processes, the selective attention paradigm (Garner, 1974; 1976) was applied in a number of studies (Schweinberger et al., 1998; Schweinberger et al., 1999; Yakel et al., 2000). By investigating the influence of variations of a task-irrelevant stimulus dimension on hypothesised perceptual processes, some researchers concluded that identity exerts an influence on the processing of facial speech (Schweinberger et al., 1998; Yakel at al., 2000) and emotional expression (Schweinberger et al., 1998; Schweinberger et al., 1999), but not vice versa (Schweinberger et al., 1998; Schweinberger et al., 1999). This "asymmetric interaction" has been interpreted in the sense of a parallel-contingent processing of facial identity, emotional expression and facial speech (Schweinberger et al., 1999), seriously challenging the notion of a strict functional encapsulation of face perception processes.

The present experiments tested whether findings of asymmetric interactions between the processing of facial dimensions (Schweinberger et al., 1998; Schweinberger et al., 1999) can be generalized across a range of stimulus characteristics and reflect a fixed architecture of face perception processes. Alternative explanations for discrepancies between studies in line with the notion of a parallel processing of identity, expression (Bruce, 1986; Etcoff, 1984; Young et al., 1986) and speechreading (Campbell et al., 1996a) and studies suggesting at least partly integrated processes (Rosenblum et al., 2002; Schweinberger et al., 1998; Schweinberger et al., 1999; Yakel et al., 2000) were considered and systematically scrutinized. Different methods were used with the intent to gather converging evidence either for or against the independence model (Bruce et al., 1986). These included the application of morphing, a digital picture editing technique, which allows for a selective manipulation of facial dimensions such as identity or expression (Beale et al., 1995; Calder et al., 1996; Schweinberger et al., 1999), the use of dynamic stimuli, the influence of face familiarity on the processing of expression and facial speech, and in particular a thorough investigation of potential pitfalls when applying the selective attention paradigm to complex stimuli such as faces.

The first part of this dissertation (Experiments 1 to 5) was dedicated to speechreading and its relationship to face identity processing. Both processes are thought to be mediated by different anatomical brain areas, which might however closely interact with one another. In the case of identity processing, regions in the inferior temporal lobe, in particular the "fusiform face area" have been suggested to be of major importance (e.g. Kanwisher et al., 1997) while speechreading has been reported to activate areas in the auditory cortex (Campbell et al., 2001) and in the superior temporal sulcus, which is thought to be involved in the processing of socially relevant stimuli (for a review see Haxby et al., 2000). So far only few studies have investigated possible influences of identity on speechreading. Moreover, experiments that presented dynamic material seem to be the exception, although it has been suggested that in particular for speechreading, dynamic information is of major importance (Rosenblum et al., 1996). Therefore, a possible influence of identity processing on speechreading (Schweinberger et al., 1998; Yakel et al., 2000) was investigated by presenting static *and* dynamic stimuli.

Experiments 1 to 3 applied a Garner type selective attention paradigm and suggested that task-irrelevant speaker variations reliably decrease speechreading speed for relatively simple vowel utterances when dynamic speech is provided and participants are unfamiliar with the presented speakers. The finding of task-irrelevant speaker variations on static faces (Schweinberger et al., 1998) could not be reliably replicated, although there was a trend in this direction. It was argued that processing costs in the orthogonal condition of the selective attention paradigm might provide evidence for speaker specific dynamic information held active in working memory, which may have to be recomputed when the speaker changes from trial to trial. Similar results are well established for the acoustic modality (Mullenix et al., 1989) and have been interpreted accordingly. The finding of speaker interferences for speechreading from dynamic stimuli is in line with reports of diminished accuracy for mixed speaker lists (Yakel et al., 2000). It is unlikely that the RT increase in the orthogonal condition of Experiment 2 was caused by an increase of task-relevant facial speech information from the control to the orthogonal condition. First, an increase of stimulus variability does not necessarily lead to a linear increase of task difficulty (Mullenix et al., 1990) so that the relative RT increase by adding stimuli might be larger for initially smaller stimulus sets. In Experiments 1 to 3, a considerable amount of overall stimulus variability was provided in the control condition. In addition, if the larger stimulus sets in the orthogonal condition were responsible for the orthogonal interference effect in Experiment 2, the same effect should have been found in Experiments 1 and 3, which used static and static-sequential faces, but applied the same design. In fact, the discrepancy between the findings of Schweinberger et al. (1998) for static faces and Experiment 1 (although there was a similar trend) might arise from differences in overall relevant information variability in the control conditions of both experiments.

The comparison of performance between static, static-sequential and moving faces also provided further evidence for a crucial role of dynamic information for speechreading (see also Rosenblum et al., 1998). The fact that no reliable interference was found for either static or static-sequential stimuli might suggest a moderating role of dynamic information on interactions between facial speech and identity processes. This possibility raised the question, whether idiosyncratic dynamic speaker information is permanently available for familiar faces and can be used in order to improve speechreading.

This was investigated in Experiments 4 and 5, which used the same paradigm as Experiment 2, but differed with respect to the group of participants. Whereas in the previous Experiment 2 all subjects had been unfamiliar with both presented speakers, participants in Experiments 4 and 5 were either familiar only with Speaker A or Speaker B. Somewhat unexpected, the effect of orthogonal interference of speaker variations disappeared, when one of the two speakers was personally familiar. This contrasts with previous findings based on static faces (Schweinberger et al., 1998). The reason for the absence of orthogonal interference for dynamic faces if one speaker is familiar is not completely clear. Possibly, idiosyncratic dynamic speaker characteristic might be stored in long-term memory, thereby decreasing the interfering influence of speaker variations by a decreased need of trial-to trial re-computations. The overall influence of familiarity on speechreading speed provides some evidence for this interpretation: the findings indicated slightly faster speechreading from highly familiar faces. Although this effect was only marginally significant, it is in line with previous research on static faces (Schweinberger et al., 1998: but see also Campbell et al., 1996a). The question, whether person specific dynamic speech characteristic can be stored and used to speechread more efficiently cannot be answered conclusively at present and deserves further investigation, perhaps using more complex speech material such as words or whole sentences.

In the context of the integrated neural model of face perception put forward by Haxby et al. (2000), the results suggest that face recognition processes, generally assumed to be mediated by structures in the inferior temporal lobe might interact with brain regions involved in the processing of facial speech. These regions are probably located in the auditory cortex of the temporal lobe and the superior temporal sulcus (STS). It has been suggested that the STS region preferably responds to dynamic input (Grossman et al., 2000), which might explain the moderating role of dynamic information on orthogonal interference of identity on speechreading found in Experiments 1 to 3. With respect to the relationship between identity processing and speechreading it can be said that the results challenge the idea of a strict functional distinction between both processes and point into the direction of an early integration of idiosyncratic speaker characteristics and especially dynamic speech information. The experiments provide some evidence for a long-term storage of speaker specific characteristics which might be used to optimise speechreading from familiar faces.

The second part of this dissertation (Experiments 6-14) looked at possible dependencies between the processing of facial identity and emotional expression. The question whether previous reports of an asymmetric interaction between the two dimensions in the selective attention paradigm (Schweinberger et al., 1998; Schweinberger et al., 1999) can be generalized across a wider range of stimulus characteristics and conditions was systematically explored in Experiments 6, 7 and 11-14. Traditionally, the Garner paradigm has been used to investigate the processing of relatively basic stimulus dimensions such as e.g. colour and form. Faces represent a much more complex type of material and this might have important consequences when applying the paradigm to face perception. Different factors which possibly affect orthogonal interference in the selective attention paradigm and therefore might restrict the validity of previous studies were taken into consideration.

In Experiments 6 and 7, identity processing was uninfluenced by task-irrelevant variations of expression. Similarly, orthogonal variation of face identity categories had no effect on classifications of expression. These results were at variance with reports of an asymmetric interaction between expression and identity processing (Schweinberger et al., 1998; Schweinberger et al., 1999). It was argued that the contrasting results might have been either due to differences in the definition of identity variations between the studies, a better control for increasing task difficulty from the control to the orthogonal condition in the expression task of the present study or to an overall difference in relative processing speed between identity and expression in previous studies (Schweinberger et al., 1998; Schweinberger et al., 1999).

The possible role of relative processing speed was explored in Experiments 11 and 12. It was argued that the more difficult of two dimensions investigated in the selective attention paradigm might be more susceptible to variations of the easier one. This deliberation was based on the observation that in studies that had found an influence of identity variations on expression analysis, the later tended to slower, while no differences in relative processing speed were present in Experiments 6 and 7. However, the pattern of results found in Experiments 6 and 7 was basically replicated in Experiments 11 and 12, where expression was overall the more difficult dimension. In comparison to the studies by Schweinberger et al. (1998; 1999), a possible increase of overall relevant facial information from the control to the orthogonal condition which might have differed between the identity and the expression tasks in the cited studies was controlled for by increasing and counterbalancing overall stimulus variability.

Another major difference between the studies refers to the definition of "identity variations". Whereas Schweinberger et al. (1998; 1999) varied individual face identity, the control condition of the present experiments allowed for individual face variations, but did so within constant familiarity categories.

Experiments 13 and 14 used a selective attention paradigm and examined the contribution of both factors to the conflicting results and also investigated interactions between the processes, when identity was more difficult to classify than expression. This was achieved by presenting similar faces without external features. In addition, the possibility that the reported asymmetric interaction between emotion and identity processing in the selective attention paradigm might have been the result of a stronger susceptibility of expression processes to unspecific pictorial variations (colour vs. b/w images) was explored. Importantly, no effect of task-irrelevant pictorial variation was present.

The other major finding of Experiments 13 and 14 was an overall condition effect for reaction times. The RT increase was only significant for the comparison of the control with an orthogonal facial condition which included task-irrelevant facial variations. Most importantly, the effect did not interact with experiment and was also significant in the expression task, if participants had not encountered the stimulus material before. In the identity task, the numerical effect decreased and in the expression task it completely disappeared for participants who had repeatedly seen the stimuli in a previous experiment. Somewhat counter-intuitively, this showed that is was easier to ignore task irrelevant variations of a dimension, if this particular dimension had been attended to in a previous experiment, strongly supporting the notion of independent processes.

These experiments contributed significantly to the clarification of the conflicting results between Experiments 6, 7, 11 and 12 and the studies of Schweinberger et al (1998; 1999). They suggest that the later findings were probably not influenced by pictorial variations but might have been influenced at least to some extent by increases in task difficulty from the control to the orthogonal condition in the expression task, but not in the identity task, as a result of the presence of external facial features. The results of Experiment 13 and 14 further implied that the increase of task-relevant information from the control to the orthogonal condition might have a stronger impact on the more difficult dimension, attributing a moderating role to task difficulty and relative processing speed.

Experiments 8 to 10 investigated whether a selective manipulation of either facial identity or emotional expression by means of morphing has an affect on classifications of the respective non-manipulated facial dimension. It was argued that a more efficient analysis of emotional expression for familiar faces would argue for an integrated processing of both dimensions. No clear-cut evidence in favour of a familiarity effect on expression classifications was found. This is basically in line with results from Experiments 6 and 7 and studies that used matching or speeded classification tasks (Bobes et al., 2000; Bruce, 1986; Campbell et al., 1996a; Young et al., 1986). The finding is also in line with the concept of two independent, parallel routes for the processing of identity and expression (Bruce at al., 1986).

However, the major finding of Experiment 8 was that recognition of familiar faces was most effective for moderately happy expressions, while classifications of unfamiliar faces were independent of expression, *as previously reported* (Schweinberger et al, 1999). This result suggests that long-term representations of familiar faces, conceptualised as FRUs (Bruce et al., 1986) might be less abstract than previously thought and seem to preserve information about typical emotional expressions. The results of Experiment 7 showed a similar interaction between familiarity and expression and represent converging evidence with a completely different stimulus set.

Overall, the results of Experiments 6 to 14 are in line with the notion of a considerable degree of functional independence between the *processing* of facial identity and facial expression. However, they provide considerable evidence that *representations* of familiar faces do not only preserve structural, but also expressive information.

For the processing of emotional expression, it has been suggested that a range of cortical and sub-cortical structures is involved, amongst others encompassing the STS region, the amygdala and the basal ganglia (see also section 1.2.3), possibly with a different role of each structure for the processing of a particular expression. This implies that theoretically, a possible *interaction between* expression and identity processing might be different for each expression. In this series of experiments, happy, angry and neutral expressions were investigated. Overall it was found that familiarity did not substantially contribute at least for the processing of these expressions. However, structures in the inferior temporal lobe that are possibly involved in the storage and retrieval of familiar faces might also use typical expressive information. This does not necessarily mean that hypothesized expression and identity processing modules in the inferior temporal lobe and the supra-temporal gyrus interact with each other or that the face recognition system has a direct access to emotional expressions. Alternatively, it is possible that emotional expressions have an influence on face recognition because expressive configurations might alter the configural information used by the face recognition system, resulting in a better representation of more typical expressions.

To summarize, thorough testing of the relationship between identity and expression processes in the selective attention paradigm suggested that the previously reported asymmetric interaction (Schweinberger et al., 1998; Schweinberger et al., 1999) does not generalize across a wide range of experimental conditions, but might be restricted to small stimulus sets where it may be influenced by differences in overall task difficulty. Overall, the experiments are compatible with the notion of a parallel and independent processing of facial identity and expression (Bruce et al., 1986) but provide evidence that FRUs preserve information about typical expressive configurations.

9 Reference List

Adolphs. R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature, 372(6507), 669-672.

Allison, T., Ginter, H., McCarthy, G., Nobre, A.C., Puce, A., Luby, M., & Spencer, D.D. (1994). Face recognition in human extrastriate cortex. Journal Of Neurophysiology, 71(2), 821-825.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends In Cognitive Sciences, 4(7), 267-278.

Allison, T., Puce, A., Spencer, D.D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cerebral Cortex, 9, 415-430.

Baudouin, J.Y., Gilibert, D., Sansone, S., & Tiberghien, G. (2000). When the smile is a cue to familiarity. Memory, 8(5), 285-292.

Baudouin, J.Y., Sansone, S., & Tiberghien, G. (2000). Recognizing expression from familiar and unfamiliar faces. Pragmatics and Cognition, 8(1), 123-146.

Beale, J.M., & Keil, F.C. (1995). Categorical effects in the perception of faces. Cognition, 57, 217-239.

Bell, C. (1844). The anatomy and philosophy of expression. (3rd Edition ed.). London: George Bell.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience, 8, 551-565.

Bentin, S., & Deouell, L.Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cognitive Neuropsychology, 17(1-3), 35-54.

Blair, R.J.R., Morris, J.S., Frith, C.D., Perrett, D.I., & Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. Brain, 122, 883-893.

Blonder, L.X., Bowers, D., & Heilman, K.M. (1991). The role of the right hemisphere in emotional communication. Brain, 114, 1115-1127.

Bobes, M.A., Martin, M., Olivares, E., & Valdes-Sosa, M. (2000). Different scalp topography of brain potentials related to expression and identity matching of faces. Cognitive Brain Research, 9(3), 249-260.

Bodamer, J. (1947). Die Prosop-Agnosie. Archiv für Psychiatrie und Nervenkrankheiten, 179, 6-53.

Bornstein, B. (1963). Prosopagnosia. In L. Halpern (Ed.), Problems of dynamic neurology. (pp. 283-318). Jerusalem: Hadassah Medical School.

Bowers, D., Bauer, R.M., Coslett, H.B., & Heilman, K.M. (1985). Processing of faces by patients with unilateral hemisphere lesions. 1. Dissociation between judgements of facial affect and facial identity. Brain and Cognition, 4(258), 272

Bowers, D., Blonder, L.X., Feinberg, T., *& Heilman, K.M. (1991).* Differential impact of right and left hemisphere lesions on facial emotion and object imagery. Brain, 114, 2593-2609.

Braida, L.D. (1991). Crossmodal integration in the identification of consonant segments. Quarterly Journal Of Experimental Psychology Section A-Human Experimental Psychology, 43(3), 647-677.

Bruce, V. (1982). Changing faces - visual and non-visual coding processes in face recognition. British Journal of Psychology, 73(FEB), 105-116.

Bruce, V. (1986). Influences of familiarity on the processing of faces. Perception, 15(4), 387-397.

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., & Linney, A. (1991). Recognizing facial surfaces. Perception, 20(6), 755-769.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. Perception, 23(7), 803-822.

Bruce, V., & Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77, 305-327.

Bruce, V., & Young, A. (1998). In the eye of the beholder. Oxford New York Tokyo: Oxford University Press.

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

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. Brain and Cognition, 2(3), 257-284.

Burt, D.M., & Perrett, D.I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. Neuropsychologia, 35(5), 685-693.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. British Journal of Psychology, 81, 361-380.

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. Memory & Cognition, 27(1), 139-151.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. Psychological Science, 11(5), 429-433.

Calder, A.J., Young, A.W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. Journal Of Experimental Psychology-Human Perception And Performance, 26(2), 527-551.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. Visual Cognition, 3(2), 81-117.

Calder, A.J., Young, A.W., Rowland, D., Perrett, D.I., Hodges, J.R., & Etcoff, N.L. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. Cognitive Neuropsychology, 13(5), 699-745.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P.K., Woodruff, P.W.R., Iverson, S.D., & David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science, 276(5312), 593-596.

Campbell, R. (1978). Asymmetries in interpreting and expressing a posed facial expression. Cortex, 14, 327-342.

Campbell, R. (1986). The lateralization of lipread sounds: a first look. Brain and Cognition, 5, 1-21.

Campbell, R. (1992). The neuropsychology of lipreading. Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences, 335(1273), 39-45.

Campbell, R., Brooks, B., deHaan, E., & Roberts, T. (1996a). Dissociating face processing skills: Decisions about lip-read speech, expression, and identity. The Quarterly Journal of Experimental Psychology, 49A(2), 295-314.

Campbell, R., & De Haan, E.H.F. (1998). Repetition priming for face speech images: Speech-reading primes face identification. British Journal of Psychology, 89, 309-323.

Campbell, R., DeGelder, B., & deHaan, E. (1996b). The lateralization of lip-reading: A second look. Neuropsychologia, 34, 1235-1240.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory visual fusion illusions - 4 case-studies and their implications. Neuropsychologia, 28(8), 787-802.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lip reading. A neurological dissociation. Brain, 109, 509-521.

Campbell, R., Zihl, J., Massaro, D., Munhall, K., & Cohen, M.M. (1997). Speechreading in the akinetopsic patient, L.M. Brain, 120, 1793-1803.

Carey, S. (1992). Becoming a face expert. Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences, 335, 95-103.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. Science, 195, 312-314.

Damasio, A.R., Tranel, D., & Damasio, H. (1989). Disorders of visual recognition. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology. Amsterdam: Elsevier Science Publishers.

Damasio, A.R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. Annual Review Of Neuroscience, 13, 89-109.

Darwin, C. (1872). The expression of the emotions in man and animals. London: John Murray.

Davies, G.M., & Milne, A. (1982). Recognizing faces in and out of context. Current Psychological Research, 2, 235-246.

DeRenzi, E. (1986). Prosopagnosia in 2 patients with CT scan evidence of damage confined to the right-hemisphere. Neuropsychologia, 24(3), 385-389.

DeRenzi, E., Faglioni, P., & Spinnler, H. (1968). The performance of patients with unilateral brain damage on face recognition tasks. Cortex, 4, 17-34.

DeRenzi, E., Peranid, D., Carlesimo, G.A., Silveri, M.C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right-hemisphere - an MRI and PET study and a review of the literature. Neuropsychologia, 32(8), 893-902.

Duchenne, d.B.G.B. (1862). Mécanisme de la physionomie humaine ou analyse électro-physiologique de l'expression des passions applicable a la pratique des arts plastiques. Paris: Renouard.

Eimas, P.D., Tartter, V.C., Miller, J.L., & Keuthen, N.J. (1978). Asymmetric dependencies in processing phonetic features. Perception & Psychophysics, 23(11), 12-20.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clinical Neurophysiology, 111(4), 694-705.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J.K.Cole (Ed.), Lincoln, Nebraska: University of Nebraska Press.

Ekman, P. (1982). Emotion in the human face. (2nd edition ed.). Cambridge: Cambridge University Press.

Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200.

Ekman, P., & Friesen, W.V. (1976). Pictures of facial affect. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W.V., & Ellsworth, P. (1982). Does the face provide accurate information? In P. Ekman (Ed.), Cambridge: Cambridge University Press.

Ellis, A.W. (1986). Processes underlying face recognition. In R. Bruyer (Ed.), The neuropsychology of face perception and emotional expression. (pp. 1-27). Hillsdale, N.J.: Erlbaum.

Ellis, A.W. (1989). Neuro-cognitive processing of faces and voices. In A. W. Young & H. D. Ellis (Eds.), Handbook of research on face processing. (pp. 207-215). Amsterdam: North-Holland:

Ellis, A.W., Young, A.W., & Flude, B.M. (1990). Repetition priming and face processing: priming occurs within the system that responds to the identity of a face. The Quarterly Journal of Experimental Psychology, 42A(3), 495-512.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception, 8, 431-439.

Endo, N., Endo, M., Kirita, T., & Maruyama, K. (1992). The effects of expression on face recognition. Tohoku Psychologia Folia, 51, 37-44.

Etcoff, N. (1984). Selective attention to facial identity and facial emotion. Neuropsychologia, 22, 281-295.

Etcoff, N.L. (1984). Perceptual and conceptual organization of facial emotions: hemispheric differences. Brain and Cognition, 3, 385-412.

Etcoff, N.L., & Magee, J.J. (1992). Categorical perception of facial expressions. Cognition, 44(3), 227-240.

Farah, M.J., & Aguirre, G.K. *(1999). Imaging visual recognition: PET and fMRI studies of the* functional anatomy of human visual recognition. Trends In Cognitive Sciences, 3(5), 179-186.

Fowler, C.A., & Dekle, D.J. (1991). Listening with eye and hand - cross-modal contributions to speech-perception. Journal of Experimental Psychology: Human Perception and Performance, 17, 816-828.

Fridja, N.H. (1993). Moods, emotion episodes, and emotion. In M. Lewis & J. M. Haviland (Eds.), Handbook of emotions. (pp. 381-404). New York: Guilford.

Garner, W.R. (1974). The processing of information and structure. Potomac, M.D.: Erlbaum.

Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice processes. Cognitive Psychology, 8, 98-123.

Gauthier, I., Behrmann, M., & Tarr, M.J. (1999). Can face recognition really be dissociated from object recognition? Journal Of Cognitive Neuroscience, 11(4), 349-370.

Gauthier, I., & Logothetis, N.K. (2000a). Is face recognition not so unique after all? Cognitive Neuropsychology, 17(1-2-3), 125-142.

Gauthier, I., & Nelson, C.A. (2001). The development of face expertise. Current Opinion in Neurobiology, 11(2), 219-224.

Gauthier, I., Tarr, M.J., Moylan, J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (2000b). Does visual subordinate-level categorisation engage the functionally defined fusiform face area? Cognitive Neuropsychology, 17(1-2-3), 143-163.

Green, K.P., & Kuhl, P.K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. Journal Of Experimental Psychology-Human Perception And Performance, 17(1), 278-288.

Green, K.P., Meltzoff, A.N., & Stevens, E.B. (1991). Integrating speech information across talkers, gender, and sensory modality - female faces and male voices in the McGurk effect. Perception & Psychophysics, 50, 524-536.

Green, K.P., Tomiak, G.R., & Kuhl, P.K. (1997). The encoding of rate and talker information during phonetic perception. Perception and Psychophysics, 59, 675-692.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. Journal of Cognitive Neuroscience, 12(5), 711-720.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. Cerebral Cortex, 10(1), 69-81.

Hancock, P.J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. Trends In Cognitive Sciences, 4 (9), 330-337.

Hasselmo, M.E., Rolls, E.T., & Baylis, G.C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual-cortex of the monkey. Behavioural Brain Research, 32(3), 203-218.

Haxby, J.V., Hoffman, E.A., & Gobbini, M.I. (2000). The distributed human neural system for face perception. Trends In Cognitive Sciences, 4(6), 223-233.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron, 22(1), 189-199.

Hay, D.C., & Young, A.W. (1982). The human face. In A. W. Ellis (Ed.), Normality and pathology in cognitive functions. (pp. 173-202). London: Academic Press.

Hoffman, E.A., & Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nature Neuroscience, 3(1), 80-84.

Hosie, J.A., Ellis, H.D., & Haig, N.D. (1988). The effect of feature displacement on the perception of well-known faces. Perception, 17, 461-474.

Humphreys, G.W., Donnelly, N., & Riddoch, M.J. (1993). Expression is computed separately from facial identity, and is computed separately for moving and static faces: neuropsychological evidence. Neuropsychologia, 31, 173-181.

Huynh, H., & Feldt, L.S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split plot designs. Journal of Educational Statistics, 1, 69-82.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, H.L., & Haxby, J.V. (1999). Distributed representation of objects in the human ventral visual pathway. Proceedings Of The National Academy Of Sciences Of The United States Of America, 96(16), 9379-9384.

Izard, C.E. (1992). Basic emotions, relations among emotions, and emotion cognition relations. Psychological Review, 99(3), 561-565.

Izard, C.E. (1993). Organizational and motivational functions of discrete emotions. In M. Lewis & J. M. Haviland (Eds.), Handbook of emotions. (pp. 631-641). New York: Guilford.

Kanwisher, N. (2000). Domain specificity in face perception. Nature Neuroscience, 3(8), 759-763.

Kanwisher, N. (2001). Neural events and perceptual awareness. Cognition, 79(1-2), 89-113.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17(11), 4302-4311.

Kirita, T., & Endo, M. (1995). Happy face advantage in recognizing facial expressions. Acta Psychologica, 89(2), 149-163.

Kirouac, G., & Doré, F.Y. (1983). Accuracy and latency of judgment of facial expressions of emotions. Perceptual and Motor Skills, 57, 683-686.

Kottoor, T.M. (1989). Recognition of faces by adults. Psychological Studies, 34, 102-105.

Kricos, P.B. (1996). Differences in visual intelligibility across talkers. In D. G. Stork & M. E. Hennecke (Eds.), Speechreading by humans and machines: Models, systems and applications. (pp. 43-53). Berlin: Springer.

Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1141.

Kurucz, J., & Feldmar, G. (1979a). Prosopo-affective agnosia as a symptom of cerebral organic disease. Journal of the American Geriatrics Society, 27, 225-230.

Kurucz, J., Feldmar, G., & Werner, W. (1979b). Prosopo-affective agnosia associated with chronic organic brain syndrome. Journal of the American Geriatrics Society, 27, 1-5.

LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., & Phelps, E.A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron, 20(5), 937-945.

Landis, T., Cummings, J.L., Christen, L., Bogen, J.E., & Imhof, H.G. (1986). Are unilateral right posterior cerebral-lesions sufficient to cause prosopagnosia - clinical and radiological findings in 6 additional patients. Cortex, 22(2), 243-252.

LeDoux, J. (1992). Emotion and the amygdala. In J. Aggleton (Ed.), The amygdala: neurobiological aspects of emotion, memory and mental dysfunction. (pp. 339-351). Wiley.

Leveroni, C.L., Seidenberg, M., Mayer, A.R., Mead, L.A., Binder, J.R., & Rao, S.M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. Journal of Neuroscience, 20(2), 878-886.

Lewis, M. (1993). The emergence of human emotions. In M. Lewis & J. M. Haviland (Eds.), Handbook of emotions. (pp. 223-236). New York: Guilford.

Ley, R.G., & Bryden, M.P. (1979). Hemispheric difference in processing emotions and faces. Brain and Language, 7(1), 127-138.

Liu, C.H., Collin, C.A., & Chaudhuri, A. (2000). Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo. Perception, 29(6), 729-743.

Massaro, D.W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Massaro, D.W., & Cohen, M.M. (1990). Perception of synthesized audible and visible speech. Psychological Science, 1 (1), 55-63.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face specific potentials generated in occipitotemporal cortex. Cerebral Cortex, 9(431), 444

McCarthy, G., Puce, A., Gore, J.C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. Journal of Cognitive Neuroscience, 9(5), 605-610.

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

McLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology, 21, 131-141.

Meltzoff, A.N., & Kuhl, P.K. (1994). Faces and Speech: Intermodal processing of biologically relevant signals in infants and adults. In D. J. Lewkowicz & R. Lickleiter (Eds.), The development of intersensory perception: Comparative perspectives. (pp. 335-369). Hillsdale, NJ: Erlbaum.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63(81), 97

Miller, G.A., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. Journal Of The Acoustical Society Of America, 27, 338-352.

Montgomery, A.A., & Jackson, P.L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. Journal Of The Acoustical Society Of America, 73(6), 2134-2144.

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(6603), 812-815.

Mullenix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics, 47(4), 379-390.

Munhall, K.G. (2001). Functional imaging during speech production. Acta Psychologia, 107, 95-117.

Münte, T.F., Brack, M., Grootheer, O., Wieringa, B.M., Matzke, M., & Johannes, S. (1998). Brain potentials reveal the timing of face identity and expression judgments. Neuroscience Research, 30(1), 25-34.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Cognitive Brain Research, 12(2), 225-231.

Natale, M., Gur, R.E., & Gur, R.C. (1983). Hemispheric asymmetries in processing emotional expressions. Neuropsychologia, 21, 555-565.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. Perception, 30(6), 755-764.

Paller, K.A., Gonsalves, B., Grabowecky, M., Bozic, V.S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. Neuroimage, 11(2), 98-110.

Parry, F.M., Young, A.W., Saul, J.S.M., & Moss, A. (1991). Dissociable face processing impairments after brain injury. Journal of Clinical and Experimental Neuropsychology, 13(4), 545-558.

Peng, C.Y.Y. (1989). Decoding facial expressions of emotion: social, cognitive and neuropsychological aspects. Dr. Phil. thesis, University of Oxford;

Phillips, M.L. (1997). A specific neural substrate for perceiving facial expressions of disgust. Nature, 389(6650), 495-498.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C., Bullmore, E.T., Brammer, M., & Gray, J.A. (1998). Neural responses to facial and vocal expressions of fear and disgust. Proceedings of the Royal Society of London B, 265(1408), 1809-1817.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H.Kellerman (Eds.), Emotion: Theory, research, and experience; Vol. 1, Theories of emotion. (pp. 3-31). New York: Academic Press.

Potter, D.D., & Parker, D.M. (1997). Dissociation of event-related potential repetition effects in judgments of face identity and expression. Journal of Psychophysiology, 11(4), 287-303.

Puce, A., Allison, T., Bentin, S., Gore, J.C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. Journal of Neuroscience, 18(6), 2188-2199.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. Cerebral Cortex, 9, 445-458.

Quirk, G.J., Armony, J.L., & LeDoux, J.E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. Neuron, 19(3), 613-624.

Rosenblum, L.D., Johnson, J.A., & Saldana, H.M. (1996b). Point-light facial displays enhance comprehension of speech in noise. Journal of Speech and Hearing Research, 39(6), 1159-1170.

Rosenblum, L.D., & Saldana, H.M. (1996a). An audiovisual test of kinematic primitives for visual speech perception. Journal Of Experimental Psychology-Human Perception And Performance, 22(2), 318-331.

Rosenblum, L.D., & Saldana, H.M. (1998). Time varying information for visual speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), Hearing by eye: Part 2, The psychology of speechreading and audiovisual speech, (pp. 61-81). Hillsdale, N.Y.: Earlbaum.

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. Perception and Psychophysics, 59, 347-357.

Rosenblum, L.D., Yakel, D.A., Baseer, N., Panchal, A., Nordarse, B.C., & Niehus, R.P. (2002). Visual speech information for face recognition. Perception and Psychophysics,

Sansone, S., & Tiberghien, G. (1996). Traitement de l'expression faciale et reconaissance: indépendance ou interaction? Psychologie Francaise, 39(4), 327-344.

Schweinberger, S.R. (1992). Funktionelle und neuroanatomische Aspekte der Prosopagnosie. Zeitschrift für Neuropsychologie, 3, 106-119.

Schweinberger, S.R. (1996). How Gorbachev primed Yeltsin: Analyses of associative priming in person recognition by means of reaction times and event- related brain potentials. Journal Of Experimental Psychology-Learning Memory And Cognition, 22(6), 1383-1407.

Schweinberger, S.R., Burton, A.M., & Kelly, S.W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. Perception & Psychophysics, 61(6), 1102-1115.

Schweinberger, S.R., Pickering, E., Jentzsch, I., Burton, A.M., & Kaufmann, J.M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Cognitive Brain Research, under revision

Schweinberger, S.R., & Sommer, W. (1991). Contributions of stimulus encoding and memory search to right hemisphere superiority in face recognition: Behavioral and electrophysiological evidence. Neuropsychologia, 29, 389-413.

Schweinberger, S.R., & Soukup, G.R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. Journal Of Experimental Psychology-Human Perception And Performance, 24(6), 1748-1765.

Schyns, P.G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. Cognition, 69(3), 243-265.

Sergent, J., Ohta, S., MacDonald, B., & Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: a PET study. Visual Cognition, 1(2/3), 349-369.

Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001). The neural correlates of person familiarity - A functional magnetic resonance imaging study with clinical implications. Brain, 124, 804-815.

Shuttleworth, E.C., Syring, V., & Allen, N. (1982). Further observations on the nature of prosopagnosia. Brain and Cognition, 1(3), 307-322.

Sprengelmeyer, R., Rausch, M., Eysel, U.T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. Proceedings Of The Royal Society Of London Series B-Biological Sciences, 265(1409), 1927-1931.

Sprengelmeyer, R., Young, A.W., Sprengelmeyer, A., Calder, A.J., Rowland, D., Perrett, D., Homberg, V., & Lange, H. (1997). Recognition of facial expressions: Selective impairment of specific emotions in Huntington's disease. Cognitive Neuropsychology, 14(6), 839-879.

Strauss, E., & Moscovitch, M. (1981). Perception of facial expressions. Brain and Language, 13(2), 308-332.

Streit, M., Ioannides, A.A., Liu, L., Wolwer, W., Dammers, J., Gross, J., Gaebel, W., & Muller-Gartner, H.W. (1999). Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. Cognitive Brain Research, 7(4), 481-491.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. Journal Of The Acoustical Society Of America, 26, 212-215.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audiovisual incompatibility in the perception of vowels. Quarterly Journal Of Experimental Psychology Section A-Human Experimental Psychology, 36(1), 51-74.

Summerfield, Q., McLeod, A., Grath, M., & Brooke, M. (1989). Lips, teeth, and the benefits of lip reading. In A. W. Young & H. D. Ellis (Eds.), Handbook of research on face processing. (pp. 223-233). Amsterdam, North-Holland:

Tarr, M.J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nature Neuroscience, 3(8), 764-769.

Tempini, M.L.G., Price, C.J., Josephs, O., Vandenberghe, R., Cappa, S.F., Kapur, N., & Frackowiak, R.S.J. (1998). The neural systems sustaining face and proper-name processing. Brain, 121, 2103-2118.

Thompson, L.A., & Lee, K.M. (1996). Information integration in cross-modal pattern recognition: An argument for acquired modularity. Acta Psychologica, 92(1), 79-104.

Tomkins, S.S. (1984). Affect theory. In K.R.Scherer & P. Ekman (Eds.), Approaches to emotion. (pp. 163-195). Hillsdale, NY: Erlbaum.

Tovee, M.J., & Cohen-Tovée, E.M. (1993). The neural substrates of face processing models: a review. Cognitive Neuropsychology, 10(6), 505-528.

Valentine, T., & Bruce, V. (1986). Recognizing familiar faces - the role of distinctiveness and familiarity. Canadian Journal Of Psychology-Revue Canadienne De Psychologie, 40(3), 300-305.

Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. Journal of Neurology Neurosurgery and Psychiatry, 71(2), 254-257.

Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. Journal of Speech and Hearing Research, 20, 130-145.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing - familiar faces and voices in the McGurk effect. Perception & Psychophysics, 57(8), 1124-1133.

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. The Journal of Neuroscience, 18(1), 411-418.

White, M. (2001). Effect of photographic negation on matching the expressions and identities of faces. Perception, 30, 969-983.

Woodworth, R.S., & Schlosberg, H. (1954). Experimental Psychology; revised edn. New York: Henry Holt.

Yakel, D.A., Rosenblum, L.D., & Fortier, M.A. (2000). Effects of talker variability on speechreading. Perception & Psychophysics, 62(7), 1405-1412.

Young, A.W. (1987). Configurational information in face perception. Perception, 16, 747-759.

Young, A.W. (1992). Face recognition impairments. Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences, 335(1273), 47-54.

Young, A.W. (1998). Applicability of the theoretical model. In Young, A.W., Face and mind. (pp. 131-159). New York: Oxford University Press Inc.

Young, A.W., Hay, D.C., & Ellis, A.W. (1985). The faces that launched a thousand slips - everyday difficulties and errors in recognizing people. British Journal of Psychology, 76(NOV), 495-523.

Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. Perception, 14(6), 737-746.

Young, A.W., McWeeny, K.H., Hay, D.C., & Ellis, A.W. (1986). Matching familiar and unfamiliar faces on identity and expression. Psychological Research, 48(2), 63-68.

Young, A.W., Newcombe, F., DeHaan, E.H.F., Small, M., & Hay, D.C. (1993). Face perception after brain injury - selective impairments affecting identity and expression. Brain, 116, 941-959.

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N.L., Seth, A., & Perrett, D.I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition, 63 (3), 271-313.

10 Appendix

Familiar faces presented in Experiments 6 and 7:

Name:	
1. Brad Pitt	21. Elvis Presley
2. John Travolta	22. Udo Jürgens
3. Harald Schmidt	23. Guildo Horn
4. Thomas Gottschalk	24. Elton John
5. Manfred Krug	25. Rex Guildo
6. Alfred Biolek	26. Reinhard Mey
7. Harald Juhnke	27. Heino
8. Arnold Schwarzenegger	28. Paul McCartney
9. Til Schweiger	29. Robbie Williams
10. Robert De Niro	30. Stefan Raab
11. Gerhard Schröder	31. Boris Becker
12. Prince Charles	32. Jürgen Klinsmann
13. Rudolf Scharping	33. Lothar Matthäus
14. Wolfgang Schäuble	34. Berti Vogts
15. Gregor Gysi	35. Franz Beckenbauer
16. Mikail Gorbachev	36. Jan Ullrich
17. Helmut Kohl	37. Mehmet Scholl
18. Joschka Fischer	38. Michael Schumacher
19. Oskar Lafontaine	39. Heinz Harald Frentzen
20. Boris Yeltzin	40. Mika Häkkinen

Familiar faces presented in Experiments 8-10✦ and 11-12*:

Name:
1. Franz Beckenbauer✦*
2. Helmut Kohl✦*
3. Gerhard Schröder✦*
4. Boris Yeltzin✦
5. Joschka Fischer✦
6. Lothar Matthäus✦*
7. Michael Schumacher✦*
8. Edmund Stoiber✦*