

Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception

Isabelle Bühlhoff & Fiona N. Newell

To cite this article: Isabelle Bühlhoff & Fiona N. Newell (2017): Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception, *Visual Cognition*, DOI: [10.1080/13506285.2017.1290729](https://doi.org/10.1080/13506285.2017.1290729)

To link to this article: <http://dx.doi.org/10.1080/13506285.2017.1290729>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Mar 2017.



Submit your article to this journal [↗](#)



Article views: 70



View related articles [↗](#)



View Crossmark data [↗](#)

Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception

Isabelle Bühlhoff^a and Fiona N. Newell^b

^aMax Planck Institute for Biological Cybernetics, Tübingen, Germany; ^bSchool of Psychology and Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland

ABSTRACT

Although faces and voices are important sources of information for person recognition, it is unclear whether these cues interact at a late stage to act as complementary, unimodal sources for person perception or whether they are integrated early on to provide a multisensory representation of a person in memory. Here we used a crossmodal associative priming paradigm to test whether unfamiliar voices which were recently paired with unfamiliar faces could subsequently prime familiarity decisions to the related faces. Based on our previous study, we also predicted that distinctive voices would enhance the recognition of faces relative to typical voices. In Experiment 1 we found that voice primes facilitated the recognition of related target faces at test relative to learned but unrelated voice primes. Furthermore, face recognition was enhanced by the distinctiveness of the paired voice primes. In contrast, we found no evidence of priming with arbitrary sounds (Experiment 2), confirming the special status of the pairing between voices and faces for person identification. In Experiment 3, we established that voice primes relative to no prime facilitated familiarity decisions to related faces. Our results suggest a strong association between newly learned voices and faces in memory. Furthermore, the distinctiveness effect found for voice primes on face recognition suggests that the quality of the voice can affect memory for faces. Our findings are discussed with regard to existing models of person perception and argue for interactions between voices and faces that converge early in a multisensory representation of persons in long-term memory.

ARTICLE HISTORY



Received 26 August 2016
Accepted 20 January 2017

KEYWORDS

Faces; voices; distinctiveness; priming; crossmodal

Although there have been significant advances made in our understanding of how both familiar and unfamiliar faces are recognized, and the cognitive and neural mechanisms supporting face perception (see Calder, Rhodes, Johnson, & Haxby, 2011), it is unclear how other sources of cross-modal information contribute to person perception. In particular, relatively little is known about how voice and face information interacts in the process of person perception. On the one hand, voices may provide an alternative route to person perception from faces, in which voice information may trigger identifying information in memory, such as the person's name and other biographical details. On the other hand, recent findings from neuroimaging and behavioural studies suggest that vocal and facial information may interact at an earlier stage, such that voice information may directly influence the representation of the face to affect person recognition.

Over the past few decades there has been a sustained level of research activity on face perception. Functional models such as the Bruce and Young model (1986), as well as its implementation in the form of a connectionist, interactive activation (IAC) model proposed by Burton, Bruce, and Johnston (1990), have provided an important theoretical framework for the understanding of person recognition and, bar some recent enhancements (e.g., Burton, Bruce, & Hancock, 1999; Calder & Young, 2005; Young & Bruce, 2011), the fundamental assumptions made in the model have mainly stood the test of time. Most significantly, converging evidence from a wide variety of approaches investigating face recognition, including psychophysical, neuroimaging and patient studies, provide validation for the model. For example, Haxby, Hoffman, and Gobbini (2000, 2002) proposed that face recognition was represented in the brain as

CONTACT Isabelle Bühlhoff  isabelle.buelthoff@tuebingen.mpg.de  Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

a distributed network of areas, but structured on a distinct core system centred around activity in the fusiform gyrus (Kanwisher, McDermott, & Chun, 1997; Sergent, Ohta, & MacDonald, 1992) as well as the occipital face area (e.g., Pitcher, Walsh, & Duchaine, 2011), to represent the invariant aspects of faces for identification. These areas are considered unimodal and predominantly visual (Casey & Newell, 2007; Kilgour & Lederman, 2002). Such a core system also incorporates a more anterior region of the brain, the superior temporal sulcus (e.g., Perrett et al., 1985), which represents the more changeable aspects of faces (e.g., lip movements in speech, facial expressions, eye gaze, etc.) for social cognition. This area is considered a multisensory, audio-visual region (e.g., Calder & Young, 2005) which facilitates inter-personal communication and recognition (Campanella & Belin, 2007).

Although these cognitive and neural models have been influential in explaining the results of many studies into person perception, some recent studies have challenged the notion of cognitive and functional specificity of face perception. On the one hand, studies on developmental prosopagnosia provide strong evidence in support of the idea of a distinct module for the perception of faces: prosopagnosic patients are typically impaired at identifying famous faces but have less difficulty identifying other aspects of a person such as their voice or gait (e.g., Duchaine & Nakayama, 2006). On the other hand, more detailed analyses of the neural basis of face recognition raise questions about the encapsulation of a face module. For example, using probabilistic tractography, Blank, Anwender, and von Kriegstein (2011) reported evidence of direct structural connections between the fusiform face area and voice-sensitive regions of the STS. Furthermore, Joassin et al. (2011) used fMRI to measure cortical activation to voices, faces and combinations of voices and faces. Their findings, that voice–face interactions result in greater activation in regions of the brain including the fusiform gyrus than either voice or face alone, are consistent with those of Blank et al. (2011). Other behavioural findings also question the idea of a unimodal face module. For example, although earlier reports suggested no evidence for priming between voices and faces of familiar individuals (Ellis, Jones, & Mosdell, 1997), others have suggested some evidence for crossmodal priming (Schweinberger, Herholz, & Stief, 1997), particularly with relatively

long delays between the presentation of a voice prime and face target (Stevenage, Hale, Morgan, & Neil, 2014).

It can be argued that a multisensory benefit on recognition for paired associations between voices and faces is a consequence of learning from the natural statistical properties of the world. For example, Barenholtz, Lewkowicz, Davidson, and Mavica (2014) reported better learning for face and voice pairings that were congruent for gender than gender-incongruent pairings and argued that this occurred as a consequence of the frequency of the gender-congruent voice and face signals co-occurring in the real world. Indeed, von Kriegstein and Giraud (2006) also demonstrated that voices are better remembered when previously paired with faces but this multisensory benefit did not generalize to other object and sound pairings. Other findings suggest that inputs from face-processing areas seem necessary for optimal recognition of speech as well as the speaker (see von Kriegstein, Kleinschmidt, & Giraud, 2006; von Kriegstein et al., 2008). Consistent with these reports, our previous findings also suggested a privileged status for associations between faces and voices during learning (Bülthoff & Newell, 2015). Specifically, we found that unfamiliar faces, which had been previously paired with distinctive voices during a learning session, were subsequently better remembered than faces learned with a typical voice, although no such benefit was found with non-vocal sounds. Furthermore, other studies showing greater than chance performance in matching unfamiliar voices to unfamiliar faces (e.g., Kamachi, Hill, Lander, & Vatikiotis-bateson; Mavica & Barenholtz, 2012; Smith, Dunn, Baguley, & Stacey, 2016a, 2016b) suggests that redundant, multisensory information cues can enhance person perception in the absence of semantic knowledge.

The functional models of face perception discussed above imply that voices provide an alternative or complementary route to person recognition from faces through a modality-specific voice recognition unit or VRU (Burton et al., 1991). For example, although familiar persons are not as readily identifiable from their voice as their face (see e.g., Brédart, Barsics, & Hanley, 2009), voice cues may provide an important source of information for identification when visual information is less reliable (e.g., Hanley & Turner, 2000), such as when a person is far away. More specifically, it is

suggested that voice information converges with face information relatively late in information processing, in a supramodal module representing all information about a person (i.e., person identity node or PIN). Although it is not clear from the model how facial information and vocal cues merge during the perception of unfamiliar persons, we can speculate how this can occur given what is known about the properties of sub-units of the model. Since the PIN is thought to be activated by information from familiar persons only (Burton et al., 1990; Hanley, 2014) and unfamiliar faces are thought to activate only the face recognition unit (FRU, see Bruce & Young, 1986), then it is plausible to suggest that voice and face information about famous individuals is integrated at the level of person perception (PIN). However, the evidence for direct connections between voice and face regions in the brain predict that even the recognition of unfamiliar faces should be influenced by paired voice information. Such evidence might challenge the notion of late multisensory activations between faces and voices during person perception or, at the very least, provide an impetus for future studies to search for the cognitive and neural loci of multisensory interactions in person recognition.

Based on our previous findings (Bülthoff & Newell, 2015), we predicted that the quality of an unfamiliar voice would enhance the subsequent recognition of a related unfamiliar face using a priming paradigm. Here we define “unfamiliar” as voices or faces that were pre-experimentally unknown, as opposed to pre-experimentally “familiar” faces or voices such as those of famous individuals and celebrities. The benefit from distinctive voices on subsequent face recognition reported in our previous study may reflect episodic memory processes rather than long-term representation of faces per se. We therefore adopted a cross-domain, associative priming paradigm to investigate whether distinctive relative to typical voices, which were previously paired with unfamiliar faces, were more likely to facilitate familiarity decisions to related faces. Associative priming refers to the facilitated access of information in memory when paired items are presented (Meyer & Schvaneveldt, 1971).

It is argued that priming offers a more sensitive way of probing the nature of representations in memory (e.g., Bruce, Burton, Carson, Hanna, & Mason, 1994). For example, in the case of object recognition, whereas episodic memory tasks reveal effects of

viewpoint or image size on object recognition, priming appears to be invariant to variations in such image properties (see e.g., Biederman & Cooper, 1991). Bruce et al. (1994) argued that since representations of faces – at least familiar faces – should be robust to changes in image properties such as lighting or view, priming paradigms are more likely to reveal the specific characteristics of the representation of faces in memory. In the past, however, face priming has mainly been discussed with regard to familiar stimuli and it is therefore unclear the extent to which priming could occur to recently learned face and voice pairs.

In view of these studies, if our earlier finding (Bülthoff & Newell, 2015) is true only for tasks tapping into episodic memory and not tasks which evoke processes related to long-term memory, then we should not expect a facilitation for voice primes, or even distinctive voices, on the recognition of a target face. On the other hand, if voices directly affect the nature of face representations, then priming should occur for paired voices and, more so, for distinctive voices. Moreover, since both faces and voices are related to person perception, then this priming effect should be specific to voices and not generalize to other sounds. We explore these hypotheses in the following three experiments.

Experiment 1

In this experiment, we wanted to test whether hearing a voice could facilitate the recognition of a related face to which it had previously been paired during a learning phase. Participants first learned to associate 24 unfamiliar faces with unfamiliar voices, by repeatedly presenting face–voice pairs during a learning session. We then tested face recognition performance using a familiarity decision (i.e., old–new) paradigm, known to promote priming (see e.g., Ellis, Young, & Flude, 1990) in which each test face stimulus was preceded by a voice prime. We predicted that face familiarity decisions would be facilitated by a voice prime provided the voice was associated with the face stimulus in memory and not if it were unrelated.

Method

Participants

Twenty-eight students (15 female) from the Karl-Eberhardt University participated in this experiment for pay. Their ages ranged from 18 to 39 years. All were

naïve as to the purpose of the study. All reported normal or corrected-to-normal vision and none reported any auditory impairments. All participants were native German speakers.

Visual and auditory stimuli

Static images of 48 clean-shaven, unfamiliar male faces were used as stimuli (see examples of the face stimuli in Figure 1a). These images were derived from 3D laser-scans of male heads (in-house database of the Max Planck Institute for Biological Cybernetics; Blanz & Vetter, 1999; Troje & Bühlhoff, 1996). Their size and the luminance were equated (Graf & Wichmann, 2002). Each image (256 × 256 pixels in size) subtended approximately 7° × 7° of visual angle at a viewing distance of about 70 cm. The face images were presented at the centre of a computer monitor against a grey background.

The voice stimuli were adapted from a set previously described (Bühlhoff & Newell, 2015) and consisted of 24 recordings of unfamiliar male and female speakers. We created two voice sets, each labelled as *distinctive* or *typical* voice sets. The typical set consisted of 12 recordings of male speakers saying the following German text in a neutral tone: “Das ist ein Foto von mir. Merke Dir bitte mein Gesicht!” (translated as: “This is a photo of me. Please remember my face”). The distinctive set consisted of 12 voice recordings (10 male and two female speakers). The meaning of the spoken content was the same across all recordings. It is important to note that no other identifying information such as names or occupation was included in the spoken content. Distinctiveness was conferred by various unique qualities of the voice. For example, the voices differed in their prosody, sex of the speaker or language spoken, or a combination of those characteristics. All auditory stimuli were presented via a set of headphones for a duration of 2–3 seconds and were grossly equated for loudness.

In order to validate these voice stimuli as distinctive and typical stimulus categories, we adopted a rating task previously described in Bühlhoff and Newell (2015). Eleven independent judges rated each voice stimulus according to how distinctive or typical they found the voice. They responded using three keys on a button box with the left button associated with *typical* (arbitrary value 1); the right button *distinctive* (arbitrary value 5) and the middle button if the voice

was *neither distinctive nor typical* (arbitrary value 3). The average ratings of 3.44 ($SE = 0.08$) and of 2.62 ($SE = 0.06$) for the distinctive and typical voice categories respectively differed significantly from each other [$t(11) = 8.90$, $p < .001$, Cohen’s $d = 2.61$].

Design and procedure

The total set of 48 faces was divided into four sets; one face set was paired with distinctive voices and the other set with typical voices, the faces of the other two sets were used as foils. The pairing of face set to voice category was counterbalanced across participants, ensuring that all faces were included in all priming conditions across the experiment. Within each set, we pseudo-randomly paired the faces with voices across participants, thus controlling for the potential effect of superior memorability for certain faces or voice–face pairings over others.

Figure 1b gives a graphic representation of the design. The experiment consisted of two phases: a learning phase followed by a cross-modal priming task. During learning, participants viewed each of the 24 face–voice pairs (12 distinctive pairs and 12 typical pairs) five times each, in random order, resulting in 120 trials. Participants were instructed to remember the faces viewed, the voices heard and their paired combinations. Each trial started with a fixation cross for 500 ms followed by the face–voice pair (the voice was heard while its paired face remained on the screen for 5 s), followed by a blank screen for 500 ms. Following each trial, participants pressed a response key to trigger the onset of the next trial. The subsequent test session was based on a priming paradigm in which participants were instructed to make a familiarity decision (i.e., “old” or “new” face) to each face image which followed a prime as fast and as accurately as possible. Each test trial began with a fixation cross (500 ms), followed by a voice prime for a duration of 2–3 s. The voice prime was immediately followed by an image of a face which remained on the screen until a response was made or for a maximum of 3 s. Response time was measured from the onset of the test face. A blank screen, presented for 500 ms, completed the trial.

As we had a limited number of voice recordings, all voices encountered during the learning session were used as primes in the test session. There were three priming conditions in the experiment as follows: a *related face* condition, in which a learned voice

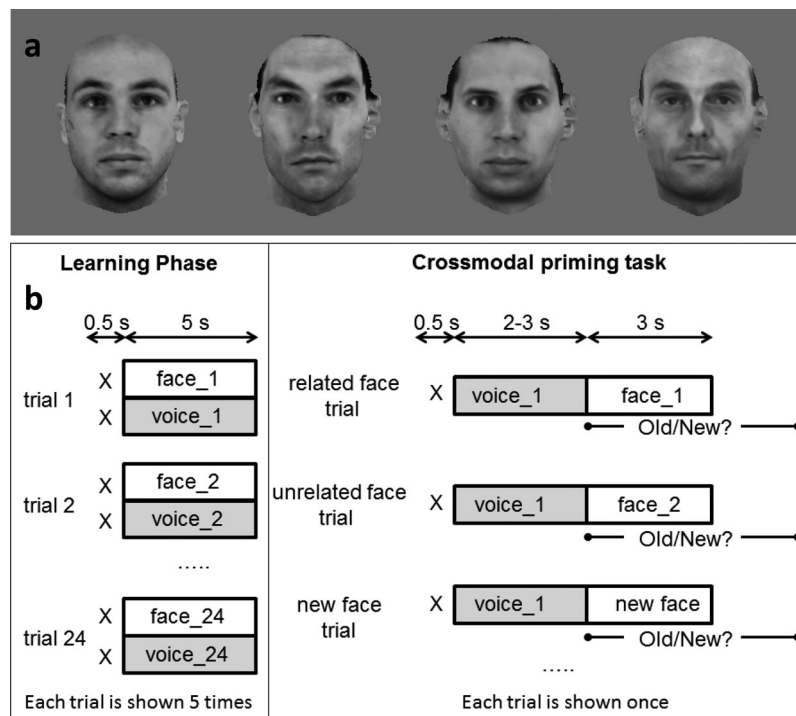


Figure 1. a. Example of face stimuli used in all experiments. b. Illustration of the main aspects of the design of Experiment 1. In Experiment 2, the same design was used, except that the voices were replaced by sounds.

primed a previously paired face during the learning session; an *unrelated face* condition, in which the learned voice primed a face that had been learned but had not been previously paired with that specific voice; a *new face* condition, in which a learned voice prime preceded a new, i.e., previously unlearned, face. The remaining 24 faces not shown during learning were used in this new face condition.

According to this design, each learned voice appeared three times as a prime and each learned face appeared twice, once in the *related face condition* and once in the *unrelated face condition* during the test phase. Each new face appeared once in the *new face condition*. Each specific voice–face pairing was presented once at test. There was a total of 72 trials (24 for each condition) in the test session. The test

session was preceded by four practice trials using stimuli not presented during the test. All trials were randomly presented across participants.

Results

We set up a mean performance accuracy threshold of 75% (calculated over all trials) for the analysis of the results. We reasoned that this threshold represented good acquisition into memory of the auditory and visual stimuli and their pairing which was required for any influence of voice priming on reaction times to be detectable. Additionally, too low performance by a participant would leave only a small number of trials available for the main response time analysis for that participant.

Five participants did not meet the required criterion, and their data were not included in further analyses. The average accuracy performance across the remaining 23 participants (13 female) over all trials was 84% and the average RT was 1359 ms. Further analyses were based on participants' mean RTs to the correct trials only and on their response sensitivity (d') (see Table 1).

A two-way, repeated measures ANOVA was conducted on the RT data with *Priming* condition and

Table 1. Mean RT and response sensitivity results for each of the voice–face priming conditions in Experiment 1. Standard error of the mean is shown in parenthesis.

Performance measure	Related face	Unrelated face	"New" faces
Response times (ms)	1337 (113)	1449 (124)	1336 (126)
Response sensitivity (d')	2.45 (.09)	2.31 (.12)	
Criterion (C)	0.36 (.05)	0.29 (.05)	
C: One-sample t -test against 0	$t(22) = 5.77$, $p < .001$	$t(22) = 7.44$, $p < .001$	

voice *Distinctiveness* as factors. We used an alpha level of .05 for all statistical tests throughout this study. A main effect of Priming condition [$F(2,44) = 3.33$, $p = .045$, $\eta^2 = .132$] was found. Pairwise comparisons, using Fisher LSD test, found that RTs to the related face were faster than to the unrelated face priming condition ($p = .029$). Response times to the unrelated face condition were also significantly longer than those to the new face condition ($p = .029$). There were no significant differences in RTs between related and new face conditions ($p = 0.96$). There was no main effect of voice *Distinctiveness* [$F(1,22) = 0.08$, $p = 0.931$]. There was a significant interaction between Priming condition and *Distinctiveness* factors [$F(2,44) = 4.93$, $p = .012$, $\eta^2 = .183$] which is plotted in [Figure 2](#). Further post-hoc analyses (Fisher LSD) revealed that this interaction was mainly due to significantly faster response times to the faces primed by *distinctive* ($M = 1262$ ms, $SE = 72$) than *typical* (1407 ms, $SE = 88$) voices in the related face priming trials only ($p = .014$). In contrast there were no significant differences in the response times between distinctive and typical voice primes in either the unrelated ($p = .099$) or new face ($p = .481$) conditions.

We were mainly interested in the response time data, however analyses of the response sensitivity (d') data and criterion C were also conducted for completeness (see [Table 1](#)). A repeated-measure ANOVA of the d' data with priming condition (related, unrelated) and voice distinctiveness (distinctive, typical) as factors revealed no main effect of priming condition [$F(2, 44) = 2.210$, $p = .151$]. There was no effect of voice distinctiveness [$F(1, 44) = 0.578$, $p = .455$] nor any clear evidence of an interaction between the factors [$F(2, 44) = 3.5252$, $p = .085$]. One-sample t -tests were conducted and revealed that the conservative criteria, C , significantly exceeded 0 (see [Table 1](#)).

Discussion

Our results suggest that recently learned voices can prime their paired faces such that familiarity decisions are facilitated for related but not unrelated faces. Furthermore, the distinctiveness of the voice primes further facilitated these response times compared to typical voices. In particular, voice distinctiveness affected familiarity decisions to the related face condition only (in which the voice primed the paired face previously learned) and not when the face was

either unrelated or new. This result supports our previous finding that the perceptual quality of a paired voice affects the recognition of a face (Bülthoff & Newell, 2015). The absence of a distinctiveness effect in the data obtained from unrelated trials further suggests that the distinctiveness of the priming voice is important for enhancing face recognition, and not necessarily any acquired distinctiveness of the learned faces through their association to distinctive voices. Albeit less interesting, we failed to find evidence for an overall facilitation in the RTs to the related trials relative to the new trials. Although this finding could be interpreted as a relative cost in response times to unrelated trials (rather than faster responses to related trials) the distinctiveness effects to the related trials suggest that response facilitation is at play. To address this point in more detail, we used a different design in Experiment 3 to confirm that related voices indeed facilitate face recognition. Participants adopted a slight response bias in which faces were more likely to be classified as “new” when they were uncertain (see Cox and Shiffrin, 2012, for a review) as indicated by their conservative criteria, C .

In sum, the results of Experiment 1 suggest that voice quality can affect subsequent recognition of a face. However, it was important to ensure that the results were specific to voice and face pairings. The following experiment was designed to test the generalizability of these cross-modal priming effects.

Experiment 2

In our next experiment we used arbitrary auditory sounds as stimuli instead of voice stimuli to test whether faces are more readily associated in memory with voices than arbitrary sounds that are not usually associated with human faces in everyday perception. We repeated the paradigm described Experiment 1 but here participants learned to associate sounds with face images during the learning session.

Method

Participants

Twenty-five participants (11 female) from the Karl-Eberhardt University of Tübingen participated in this experiment for pay. Participant ages ranged from 18

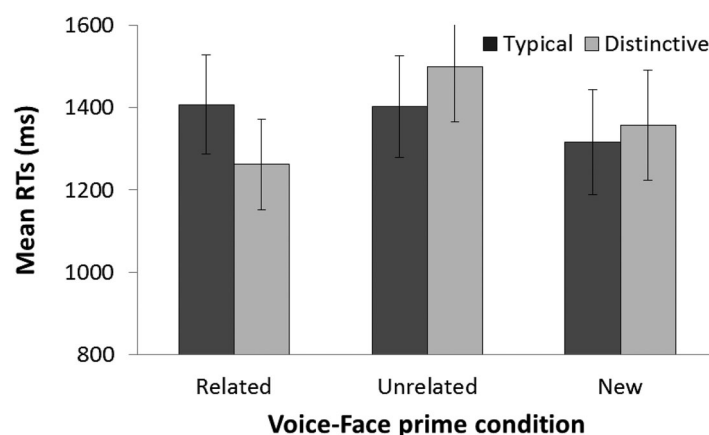


Figure 2. Plot showing the mean response times in Experiment 1. For each related, unrelated and new face priming condition, the mean response time (in ms) for trials with distinctive and typical voice primes are shown. Error bars represent ± 1 standard error of the mean.

years to 32 years. None of the participants reported any auditory or visual impairments and all were naïve to the purpose of the experiment.

Face and auditory stimuli

The same face images were used as in Experiment 1. All auditory stimuli were created using a synthesizer (Triton LE from Korg Inc.) and have been described in detail elsewhere (Bülthoff & Newell, 2015). All 12 sounds in the *typical* set consisted of different piano chords while other musical instruments and various electronic sounds were used to create the set of *distinctive* sounds. Each sound stimulus was presented for 3–5 s and all were grossly equated for loudness. Both sets differed significantly in distinctiveness as assessed by a rating test previously reported (Bülthoff & Newell, 2015). Participants were presented with the auditory stimuli via a set of headphones. Again, response time was measured from the onset of the test face.

Design and procedure

The experimental design and procedure were the same to those described in Experiment 1 (see illustration of the design in Figure 1b).

Results

The performance of five participants tested did not reach the criterion threshold of 75% correct (calculated over all trials), and a further two encountered problems during the test phase. The average correct performance of the remaining 19 (11 female)

participants was 83% (calculated over all trials), the average RT across all participants and conditions was 1241 ms. Further analyses were based on participants' mean RTs to the correct trials only and on their response sensitivity (d') (see Table 2).

Response times are shown for each of the priming conditions and *distinctive* and *typical* primes in Figure 3. A two-way, repeated measures ANOVA was conducted on the RT data with *Priming* condition (related, unrelated and new face) and sound *Distinctiveness* (typical, distinctive) as factors. This analysis revealed no effect of Priming condition [$F(2,36) = 0.33, p = .613$]. Thus, learned sounds did not significantly facilitate familiarity decisions to their related faces relative to any other condition. There was no effect of sound Distinctiveness [$F(1,18) = 1.15, p = .298$] and no interaction between the factors [$F(2,36) = 0.31, p = .733$].

As in Experiment 1, analyses of the d' data and criteria C were conducted for completeness (see Table 2). A repeated-measure ANOVA of the d' data with *Priming* condition and sound *Distinctiveness* as factors revealed no main effect of Priming condition

Table 2. Mean RT and response sensitivity results for each of the sound-face priming conditions in Experiment 2. Standard error of the mean is shown in parenthesis.

Performance measure	Related	Unrelated	New faces
Response times (ms)	1186 (103)	1217 (98)	1234 (116)
Response sensitivity (d')	2.78 (.14)	2.68 (.11)	
Criterion (C)	.21 (.07)	.26 (.05)	
C: One-sample t -test against 0	$t(18) = 3.15, p = .006$	$t(18) = 4.95, p < .001$	

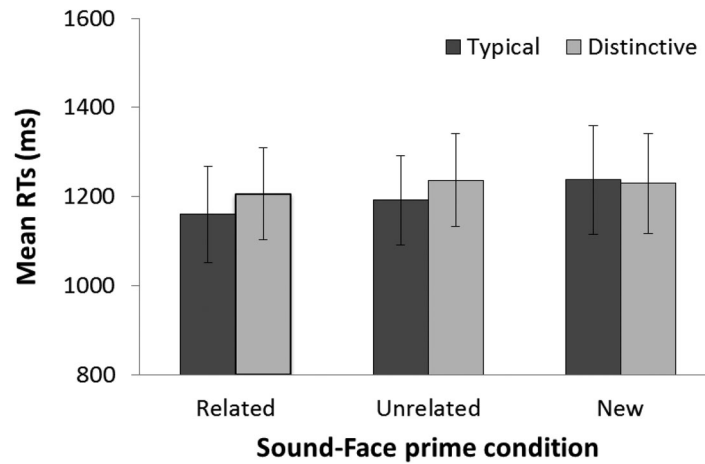


Figure 3. Plot showing the mean response times in Experiment 2. For each related, unrelated and new sound-face priming condition, the mean response time (in ms) for trials with distinctive and typical voice primes are shown. Error bars represent ± 1 standard error of the mean.

[$F(2,36) = 0.342$, $p = .566$]. There was no effect of Distinctiveness [$F(1,18) = 1.86$, $p = .190$] and no interaction between the factors [$F(2,36) = 0.005$, $p = .947$]. One-sample t -tests were conducted and revealed that the conservative criteria, C , significantly exceeded 0 (see Table 2).

Discussion

Experiment 2 was conducted to determine whether any implicit associations could emerge from learned associations between arbitrary sounds and face images, under the same learning conditions as in Experiment 1. In the previous experiment we found that unfamiliar voices that had previously been paired with unfamiliar faces in a learning session could subsequently prime familiarity decisions to their related faces. In contrast, in the present experiment we found no evidence that arbitrary sounds could prime their paired face images. This result is surprising given that the learning procedures across these experiments remained the same, yet significant priming effects to faces were found only for learned voices but not sounds.

The d' results with sound primes did not reveal any benefit for related over unrelated faces. Again, participants adopted a slight response bias in which faces were more likely to be classified as “new” when they were uncertain (see Cox and Shiffrin, 2012, for a review) as indicated by their conservative criteria, C . Since our main concern was the comparison of accuracy performance and response times between

related and unrelated face conditions, the new faces merely acted as “fillers” and the responses to these trials were not considered.

Comparing face priming with voices and sounds

To confirm that voices and sounds have different priming effects on familiarity decisions to previously learned faces only, we conducted a $2 \times 2 \times 2$ mixed-design ANOVA, with *Experiment* (Exp1, Exp2) as the between-subject factor, and both *Priming condition* (related, unrelated face) and *Distinctiveness* (distinctive, typical) as within-subject factors. See Table 3 for details of the mean response times results in each condition.

This analysis revealed a main effect of Priming condition [$F(1,40) = 7.45$, $p = .009$, $\eta^2 = .157$], with faster response times to related ($M = 1265$ ms) over unrelated ($M = 1344$ ms) face priming conditions. There was no evidence for effects of Experiment [$F(1,40) = 1.53$, $p = .223$] or Distinctiveness [$F(1,40) = 0.14$, $p = .714$]. There was a significant interaction between Priming condition and Distinctiveness [$F(1,40) = 4.38$, $p = .043$,

Table 3. Mean response times to each of the voice and sound priming conditions, to each of the distinctive and typical primes, across Experiments 1 and 2 respectively. Standard error of the mean is shown in parenthesis.

Face priming condition	Voice primes (Exp 1)		Sound primes (Exp 2)	
	Distinctive	Typical	Distinctive	Typical
Related	1262 (111)	1407 (121)	1206 (103)	1160 (108)
Unrelated	1499 (133)	1403 (124)	1237 (104)	1192 (100)

$\eta^2 = .099$] and a significant three-way interaction between Priming condition, Distinctiveness and Experiment [$F(1,40) = 4.43, p = .042, \eta^2 = .100$]. Post-hoc analyses (Fisher LSD) on the three-way interaction revealed significantly faster response times to the distinctive relative to the typical voice primes in the related face condition of Experiment 1 only ($p = .012$) but no such difference between distinctive and typical voice primes for the unrelated face condition in Experiment 1 ($p = .087$) or between the distinctive and typical sound primes in either the related or unrelated face conditions in Experiment 2 ($p = .449$ and $p = .457$ respectively). The responses to the unrelated face conditions were faster to both the typical and distinctive sounds than the typical or distinctive voices, suggesting that there may have been some interference on face recognition from unrelated voices compared to unrelated sound primes. This result further suggests that, even following learning, faces were not associated with sounds in Experiment 2.

The results of this analysis confirm that the quality of a voice, but not necessarily a sound, can affect the subsequent recognition of a previously paired face. The differences in results across the experiments suggest that the acquisition of associations between voices and faces is rapid (i.e., following just five exposures to each voice–face pair). In contrast, arbitrary sounds do not seem to share the same accessibility to representations of faces in memory. Although our results suggest that sounds neither enhanced the explicit recognition of associated faces (Bülthoff & Newell, 2015) nor implicitly facilitated the recognition of faces in the present study, we do not claim that a close association between faces and arbitrary sounds is not possible. For example, a longer learning session might result in a facilitation of related over unrelated faces in the sound priming task. However, a slight trend in our data (Figure 3, Table 3) suggests that distinctive sounds may even hinder any priming effects: unlike for voice primes, distinctive sound primes resulted in slower familiarity decisions to related faces than typical sound primes (i.e., related face priming condition). This finding confirms an observation made in our previous study of a trend for distinctive sounds to hinder recognition of their paired faces compared to typical sounds (Bülthoff & Newell, 2015).

It is interesting to note that, in both experiments, we observed evidence of a conservative response

bias to classify faces as new. Therefore, this bias seems to be independent of the prime properties (i.e., whether the prime is a voice or a sound). Alternatively, this result may hint at differences in processing between unknown and recently learned faces. However, it is beyond the scope of this present study to speculate on effects due to the degree of face familiarity but we acknowledge that this is an important issue for future research.

Experiment 3

Although the results of the previous experiments suggest that voice primes facilitated familiarity decisions to related over unrelated faces, the following experiment was conducted in order to address two potential concerns. First, the accuracy results of both previous experiments suggest that participants found it relatively difficult to learn the stimuli: the accuracy level to the “old” faces was less than 80% and lower than accuracy to the “new” faces at test. We therefore wished to test whether our findings could generalize beyond these specific task demands and, to that end, we included a learning session with more repetitions of each face–voice pair (seven instead of five) and fewer face–voice pairings (20 instead of 24) compared to Experiment 1. Our intention was, that by repeating the learning trials more often than in the previous experiment, this would allow participants to achieve greater accuracy in their performance at test. The second concern we wished to address was to ensure that the presence of an associated voice prime selectively facilitated the subsequent recognition of an associated, related face. In Experiment 1 we compared priming performance across related and unrelated face conditions and found facilitation to related faces only. Therefore, to control for the possibility of a deleterious surprise effect when an unexpected learned face followed a learned voice, in the following experiment we included both a related face priming condition and a no priming condition (i.e., when no voice was presented before the test face).

Method

Participants

In view of the subtle differences in performance obtained between conditions in the previous

experiments, a larger number of participants (and consequently larger data set) was recruited. Thirty-one participants (15 female) aged 19 years to 43 years old performed this experiment. None reported any visual or hearing impairments and none had participated in the previous experiments.

Visual and auditory stimuli

A subset of each of the face and voice stimuli used in Experiment 1 was used. The distinctiveness ratings to these sets of 10 typical (2.57 , $SE = 0.06$) and 10 distinctive voices (3.56 , $SE = 0.06$) differed significantly from each other [$t(9) = 50.40$, $p < .001$, *Cohen's* $d = 4.94$].

Design and procedure

The experiment was based on a fully factorial within subjects design with *Priming* conditions (no voice prime, voice prime) and *Face conditions* (related or new) as factors. A graphical representation of the experimental protocol, which was largely based on that from Experiment 1, is provided in [Figure 4](#).

During the learning phase participants had to associate 10 distinctive and 10 typical voices with unfamiliar faces. Each voice–face pair was repeated seven times. Those values had been determined from a pilot study in which we varied the number of test faces and the number of repetitions with the goal to obtain above-threshold responses from most participants more reliably. The order of all learning trials was random across participants. Two groups of participants learned different voice–face pairings.

The allocation of faces to each face condition (related, new) and priming condition (prime, no prime) was counterbalanced across participant groups. During the test phase, there were 40 test trials in total, with each learned face presented only once across 20 trials, and 20 new face stimuli were presented across the other 20 trials. In comparison to Experiment 1, learned faces were preceded either by their paired voice prime (consistent with the related condition in Experiment 1) or by no prime. Similarly, new faces were primed either by a learned voice or no prime was presented. Performance in the *no prime* trials served as baseline for testing the priming effect of voices for their paired, related faces. All trials with new faces served as fillers. There was no unrelated face condition for learned faces.

When related (learned) faces were presented during the test, voice *Distinctiveness* (distinctive, typical) during learning was a third factor. For all trials with a voice prime, a voice (*typical* or *distinctive*) was followed by its paired face (five trials in the typical condition and five trials in the distinctive condition) or by a new face (five trials with a distinctive prime and five trials with a typical prime). Because of the limited number of recorded voices, an experimental design was chosen that did not require new voices to be used in the test session.

The design of the test phase allowed us to directly assess the effect of a voice prime on recognition performance of learned faces in comparison to trials without voice primes. In *primed trials*, participants first heard a voice prime accompanied by the written word “Voice” on the computer monitor. For *no prime trials*, the words “No Voice” were presented and no voice was heard. In both cases, a fixation cross immediately replaced the written words for 500 ms, to ensure that the participant was alerted to the subsequent exposure of the face image in both the prime and no-prime conditions. On the basis of predictions from the IAC model (Burton et al., 1990) and previous reports about the short duration of cross-domain priming effects (see e.g., Calder, Young, Benson, & Perrett, 1996; Ellis, Young, Flude, & Hay, 1987) we were not concerned that this 500 ms interval would affect priming. The fixation cross was then followed by a test face for 2 s. Again, response time was measured from the onset of the test face. A blank screen presented for 1 s completed the trial. The participant’s task was to make a familiarity decision on each face (i.e., decide if it was old or new) as accurately and as fast as possible. Each test session was preceded by four practice trials with stimuli not used during the test. In all other ways the procedure and participants’ instructions were the same as in Experiment 1, except that that participants were informed that responses were not included if none were provided within 3 s.¹ They knew that in case they did not enter their response on time, the next trial would be triggered automatically after an audible warning beep. We included this instruction in an attempt to reduce large intra-individual variability in the response times.

¹Participants failed to provide a response within this time limit in less than 0.5% of all trials.

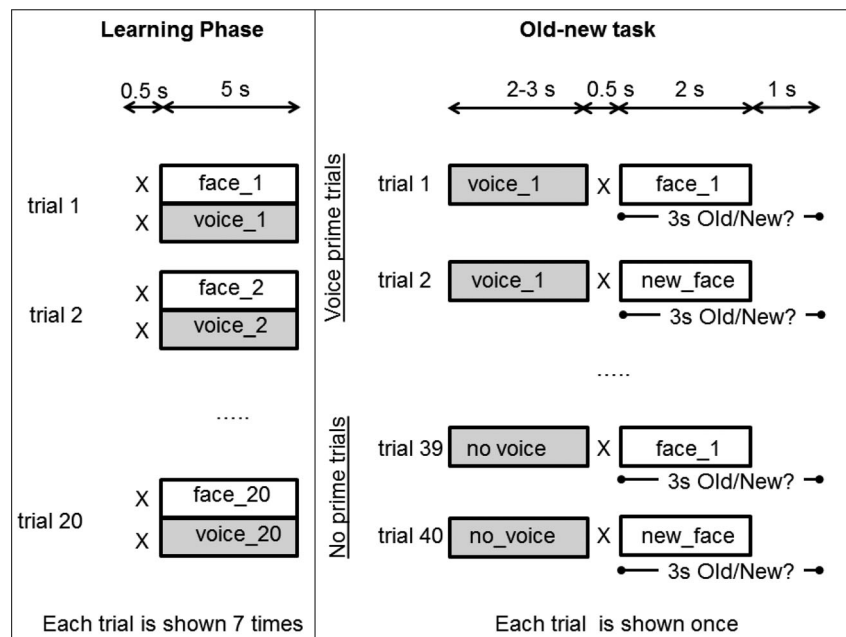


Figure 4. Main aspects of the design of Experiment 3.

Results

One participant encountered technical problems during the test phase and their data were not included in further analyses. We analysed the results of the remaining 30 participants (15 female), each of whom achieved an overall accuracy performance of over 75%. Their average accuracy performance calculated over all trials was 92% and their average RT was 1044 ms. The mean response times and response sensitivity (d') to each of the priming conditions for the learned faces and the mean performance to the new or unlearned faces are presented in Table 4.

For each participant the mean RTs were calculated from their correct trials only. Overall, participants responded slightly more rapidly to new than related faces ($M = 1013$ ms and $M = 1046$ ms, respectively),

Table 4. Mean RT and accuracy results for each of the voice prime and no-prime conditions to the related and new face conditions in Experiment 3. Standard error of the mean is shown in parenthesis.

Response measure	Related faces		New faces	
	Voice prime	No voice prime	Voice prime	No voice prime
Response times (ms)	1003 (.46)	1082 (.46)	1013 (.54)	1013 (.50)
Response sensitivity (d')	2.73 (.09)	2.61 (.10)		
Criterion (C)	0.10 (.04)	0.15 (.05)		
C: One-sample t -test against 0	$t(29) = 1.80$, $p = .082$	$t(29) = 3.38$, $p = .002$		

as well as to the *prime* than to *no prime* trials although these differences all failed to reach significance (paired t -test: all $ps > 0.062$).

To investigate whether the presence and distinctiveness of the voice prime affected subsequent familiarity decisions to the target faces, we analysed the mean response times to each of the prime conditions for the learned, related faces only. These values are shown in Figure 5. We conducted a two-way, repeated measures ANOVA with *Prime* condition (prime, no prime) and voice *Distinctiveness* (distinctive, typical) as within-subject factors. For the condition voice *Distinctiveness*, the faces appearing without any voice prime at test can still be categorized as typical or distinctive as a consequence of their previous pairing with a distinctive or typical voice during learning.

This analysis revealed a main effect of Priming condition [$F(1, 29) = 7.98$, $p = .008$, $\eta^2 = .216$], with faster response times to the voice prime over the no voice prime conditions. A main effect of the voice *Distinctiveness* [$F(1, 29) = 4.65$, $p = 0.039$, $\eta^2 = .138$] was also found, with faster response times to faces previously learned with distinctive than typical voices. There was a greater benefit on the speed of the response times for distinctive voices over typical voices in the voice prime trials (85 ms faster to distinctive voices) than in the no voice primes trials (42 ms faster to distinctive voices), as shown in Figure 5. However, the interaction between these factors failed to reach

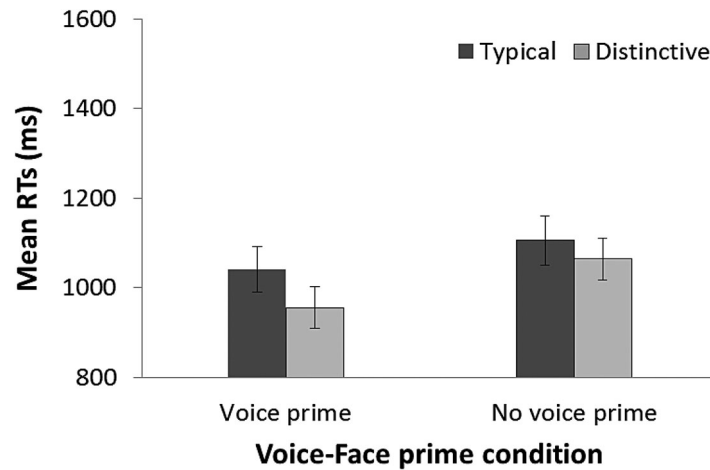


Figure 5. Plot showing the mean response times in Experiment 3. For each voice prime and no voice prime condition, the mean response time (in ms) for trials with distinctive and typical voice primes are shown. Error bars represent ± 1 standard error of the mean.

significance [$F(1,29) = 0.74$, $p = .396$]. This finding suggests that distinctive voices, presented with faces during the learning condition, benefitted the subsequent recognition of faces (i.e., the distinctiveness advantage) irrespective of whether the voice was presented with its related face during recognition (i.e., the no prime condition).

We also analysed the response sensitivity (d') data and criteria C for completeness (see Table 4). The d' data were submitted to a two-way repeated measures ANOVA with *Priming* condition (prime, no prime) and voice *Distinctiveness* during learning (distinctive, typical) as within-subject factors. This analysis revealed no effects of Priming condition [$F(1,29) = 1.19$, $p = .284$] or voice Distinctiveness [$F(1,29) = 0.015$, $p = .938$] and no interaction between these factors [$F(1,29) = 0.56$, $p = .460$]. One-sample t -tests were conducted and revealed that the conservative criteria, C , significantly exceeded 0 (see Table 4).

Discussion

The main goals of this experiment were to test whether voice distinctiveness facilitates the recognition of a previously learned face when the task was relatively less difficult, and also to test the extent to which a voice prime facilitated the recognition of a face relative to no prime condition. With regard to the first aim, the results of Experiments 1 and 2 suggested that voices but not arbitrary sounds could prime the recognition of previously paired unfamiliar faces. However the task involving

sound primes appeared to be relatively more easy (mean d' was 2.78 and 2.68 for related and unrelated trials) than that with voice primes (mean d' was 2.45 and 2.31 for related and unrelated trials), suggesting that priming per se may be affected by task demands. Here participants were provided with more repeats of the learning trials and fewer voice-face pairs to learn, and subsequent accuracy improved in the test trials relative to Experiment 1 (mean d' was 2.73 for related trials and 2.61 for trials without any voice primes). Importantly, despite the change in task demands, the results of Experiment 3 concur with those of Experiment 1 in that the distinctiveness of the voice, previously learned with a face, facilitated the subsequent recognition of that face relative to typical voices. Furthermore, voice distinctiveness improved face perception irrespective of whether the voice prime was presented or not.

Our analysis also shows a clear beneficial effect of voice priming on the speed at which a subsequent learned face target was recognized compared to trials showing learned faces without a voice prime. This finding indicates that a close association between the voice and face stimuli had been established during learning so that hearing a voice could speed up the subsequent recognition of its paired face. Furthermore, familiarity decisions to related faces in the test trials were significantly faster for distinctive than typical faces, irrespective of whether or not a voice prime was heard. We cannot, however, conclude that this is evidence for a stronger

(or more effective) priming effect for distinctive than typical voices, because of the absence of an interaction between both factors (Trial Type and Distinctiveness). However, this finding argues for a speeded up retrieval from memory of faces previously paired with distinctive relative to typical voices in addition to the facilitation from a voice prime. Finally, the results in this experiment confirm that the greater facilitation for distinctive than typical voice primes observed in Experiment 1 is due, at least in part, to the priming effect of the voices. In that first experiment, in the related condition, familiarity decisions to faces that were primed by a distinctive voice could have been facilitated not because of the prime, but because of face distinctiveness acquired cross-modally during the learning phase (Bülthoff & Newell, 2015). In this third experiment, the faster response times to the conditions in “prime” than in “no prime” trials ensure that these facilitatory effects were, at least in part, directly resulting from the priming effect of the voice.

In the absence of voice primes, our data revealed that faces that had been paired with a distinctive voice during learning were recognized faster than faces paired to typical voices. This result is somewhat consistent with the finding of our previous study (Bülthoff & Newell, 2015) in which we tested the explicit recognition of faces using an old–new recognition paradigm (i.e., without any crossmodal priming) after a shorter learning session. In that previous study, we found that faces previously paired with distinctive voices during learning were classified more accurately than faces paired with typical voices, while we found no effect in response times. In the present study we tried to ensure that the learning session resulted in high recognition accuracy at test in order to measure priming effects on response times. Under these conditions, participants performed equally well across all conditions in terms of response sensitivity, albeit with a slight tendency to classify faces as new (conservative criteria, C), but the quality of the voice during learning affected subsequent recognition speed at test. Our results are therefore consistent with the idea that self-priming (Calder et al., 1996) can occur across modalities to facilitate face perception.

Correlations between voice distinctiveness and participants' performance in Experiments 1 and 3

The results of Experiments 1 and 3 suggest that distinctive voice primes facilitated familiarity decisions to related faces relative to typical voice primes. We wondered whether a correlation could be found between the perceived distinctiveness of the voices (as assessed by the distinctiveness ratings) and participants' performance to the trials in the related face conditions in both Experiments 1 and 3. We found a weak but significant correlation between the rated distinctiveness of the voice primes and response accuracy in Experiment 1 (Pearson correlation .372, one-sided $p = .037$); no significant correlations were found for the response times. For Experiment 3, we found only a trend for a correlation between distinctiveness ratings to the voice primes and response times (Pearson correlation $-.365$, one-sided $p = .057$); no significant correlations were found for the accuracy data.² We do not have any explanation as to why a correlation is found in response time or in accuracy data only. However, in view of those results, we can merely state that the correlations point in the expected direction, whilst acknowledging that this evidence is rather weak. This finding is consistent with relatively weak effects of voice relative to face in person perception in general (e.g., Stevenage et al., 2013), but particularly with regard to crossmodal priming effects (Stevenage et al., 2014). On that note, we might expect that the distinctiveness of a face may have a more robust effect on priming of paired voices.

General discussion

Our first finding from these studies is that voice primes, which were previously learned with unfamiliar faces, subsequently facilitated familiarity decisions to their paired faces, but not to unrelated faces (i.e., faces previously learned with other voices). Moreover, the benefit for crossmodal facilitation on face recognition was specific to voice primes, and did not generalize to learned arbitrary sounds. This result is consistent with previous reports of enhanced person recognition from exposure to unfamiliar voice–face pairings (e.g., Bülthoff & Newell, 2015; Smith and

²For completeness, Pearson correlations were calculated for unrelated trials in Experiment 1 (there were no unrelated trials to analyse in Experiment 3). Both correlations, calculated for response times and accuracy data, were non-significant.

colleagues, 2016a, 2016b), as well as neuroimaging findings for direct connections between cortical areas subserving voice and face perception (e.g., Blank et al., 2011; von Kriegstein, Kleinschmidt, & Giraud, 2006). The relative ease of associating voices with faces may be due to the lifetime of experience in being exposed to bimodal stimulation of faces and voices that may subsequently allow voices to have privileged access to faces during learning that then enhances person identification (e.g., Barenholtz et al. 2014; von Kriegstein et al., 2008). Recent studies suggest that even unfamiliar voices and faces can be relatively easily matched (Kamachi, Hill, Lander, & Vatikiotis-bateson, 2003; Mavica & Barenholtz, 2012; Schweinberger, Robertson, & Kaufmann, 2007; Smith et al., 2016a, 2016b; von Kriegstein & Giraud, 2006), suggesting that these cues may be integrated without semantic knowledge of the person. Note that the difference between the results obtained in Experiment 1 and Experiment 2 cannot be explained by the existence of identity-based, cross-modal redundancies shared solely between faces and voices and not between faces and sounds as in our design we randomly paired faces to voices. In view of the findings of Smith and colleagues (2016a), which demonstrated that faces and voices in natural face–voice pairings share overlapping identity cues, future studies using natural face–voice pairings would allow for an investigation of more ecological pairings on this distinctiveness issue.

Crossmodal priming is generally indicative of a close association between the information presented in both modalities, such that information from one modality can modulate the representation in the other modality. This information may either be associated directly, in an information-driven manner (e.g., through shared, redundant information) that is reinforced through brief learning (e.g., von Kriegstein et al., 2008), or indirectly through feedback via a common resource (e.g., Naci, Taylor, Cusack, & Tyler, 2012; ten Oever et al., 2016). For example, based on the results from a neuroimaging study, Joassin et al. (2011) reported that interactions between voices and faces (static images) is underpinned by activation in supramodal regions of the brain, such as the angular gyrus and hippocampus, that can influence processing in unimodal regions of the brain associated with face and voice perception. It is interesting to speculate, however, at what point during information processing

for person recognition these interactions between faces and voices arise. With reference to the Bruce and Young (1986) model, Burton et al. (1990) argued that voices and faces should interact only at the level of the person identity node or PIN, at which point familiarity decisions are made. In other words, it is assumed that considerable processing occurs in each unimodal system (i.e., in the FRU and VRU for faces and voices respectively) prior to convergence at the PIN. Furthermore, it is assumed that the PIN stores abstract representations of familiar faces that are robust to incidental changes that can occur with, for example, different lighting or viewpoint conditions. Once the PIN is activated this then allows for access to other personal information including the person's name and other semantic details (Bruce & Young, 1986; Burton et al., 1990). However, these cognitive models of face recognition were largely developed on the basis of the recognition of familiar persons and it is therefore unclear how they can account for the perception of relatively unfamiliar faces (Hancock, Bruce, & Burton, 2000), or indeed how the learning of novel face and voice pairs affect the nature of these interactions.

Although our study tested the effects of newly learned, rather than entirely unfamiliar, faces and voices we assert that these newly learned faces are more likely to activate the earlier FRU only (Bruce & Young, 1986; Young & Bruce, 2011), as image-based representations (see Johnston & Edmonds, 2009, for a review) rather than the PIN. We can make this assertion for two reasons. First, the procedure adopted in the experiments reported here, that is, of repeatedly showing the same face image face during learning (five or seven times in Experiments 1 and 3 respectively) in the absence of any other identity information (name or semantic details such as occupation) is likely to result in a more pictorial or low-level perceptual description of the face in memory rather than a rich, abstract representations of each individual. Second, the properties of the FRU and PIN (Bruce & Young, 1986, 2011; Burton et al., 1990) make it more parsimonious to assume that only the FRU was activated. For example, it has been argued that the PIN stores identity-specific, semantic information that can be accessed via multisensory information from other nodes such as the FRU and VRU, which each contain rich unimodal, information about a person. More specifically, representations at the level of the PIN

are thought to be robust to incidental changes to the percept (such as changes in face viewpoint or voice pitch) to allow for identity decisions to be readily made. The FRU is activated by any view of a face that is processed at earlier stages in the functional model as a structural code (Bruce & Young, 1986). Some studies have suggested differences in the neural representation of faces that are visually familiar only (e.g., images from magazines), versus those that are unknown, with greater activation to unfamiliar faces in early face regions of the cortex (e.g., Rossion, Schiltz, Robaye, Pirenne, & Crommelinck, 2001), although others reported no differences in face familiarity in face-selective regions (see Natu and O'Toole, 2011, for a review). Therefore, it is not sufficiently clear at what level repeated face images are processed during face perception, but it seems unlikely that repeated exposure to the same image of a (previously unfamiliar) face is sufficient to activate the PIN and more likely that activation occurs earlier in information processing such as the level of the FRU. Interestingly, Bruce and Young (1986) originally proposed that familiarity decisions are made at the level of the FRU (and, similarly, voice familiarity at the level of the VRU). If it is the case that the FRU is the node most likely to be activated by unfamiliar faces, then we can speculate that our results suggest crossmodal influences occur earlier in face processing than previously thought, possibly directly between the FRU and VRU nodes. Although our data do not speak directly to the issue of where these interactions actually occur within the Bruce and Young functional model specifically, at the very least, any model of face perception needs to take into account our findings that voices can directly prime the recognition of unfamiliar faces. Moreover, it is noteworthy that this priming occurred in the absence of any other personal knowledge of the faces, such as their names or occupation, that could provide a possible route to the PIN to account for the crossmodal priming effects (such as those reported in Barsics and Brédart, 2012).

The interactive activation model of Burton and colleagues (1990) may propose that, in the paradigm of Experiment 1, the voice PIN is competing with the PIN for the subsequent face identity hence slowing down the response found for faces primed by unrelated voice primes. The same argument might be expressed for competition between identity information at the level of the FRUs. In both cases, this

competition might be reflected in slower response times and reduced accuracy performance as we observed in this experiment for faces primed by unpaired voices compared to faces primed by their paired voices, thus it leaves the issue of where the interaction might happen open.

Our second main finding is that the perceptual quality of a voice, in other words its distinctiveness, can modulate the recognition of a paired face. In both Experiments 1 and 3, we found that response times to faces primed by distinctive voices were, on average, 115 ms faster than faces primed by typical voices. It is important to note that the effect of distinctive voices on facilitating familiarity decisions to faces was specific only to faces previously paired with the voices and not learned faces that were unrelated to those voices. One interesting finding from Experiment 1 was that there was no effect of voice distinctiveness on response times to previously unlearned ("new") faces. This result suggests that there was nothing inherent in a distinctive voice stimulus which could act as an arousing stimulus or attentional cue to the current goals of the task as previously reported (e.g., Manly et al., 2004).

The results of Experiment 3 suggested that voice primes were necessary for the largest effect of facilitation from distinctiveness on familiarity decisions to a target face to occur. Burton et al. (1990) referred to this effect as "self-priming" and made a similar prediction that familiarity decisions to a person's name would be facilitated if it was immediately preceded by an image of that person's face. This effect was considered time-sensitive with cross-modal facilitation occurring only with short lags between the prime and target. Calder and Young (1996) provided further insights into "self-priming" and found that repeated exposure to target stimuli (names in this case) did not affect the facilitation from either within-modal (repeated name) or cross-modal (the person's face) primes. Furthermore, Calder, Young, Benson, and Perrett (1996) reported greater effects of self-priming on familiarity decisions to target names for distinctive than typical faces. Most pertinent to the results of our study, they argued that these effects arise due to distinctive faces producing a "low level of inhibition within the FRU pool, allowing the prime's FRU, and consequently its PIN, to rapidly reach a high level of activation" (Calder et al., 1996, p. 158). Based on their conclusion, our findings for a

larger facilitation on familiarity decisions to target faces with distinctive than typical voice primes suggest that the low level of inhibition within the FRU occurs due to direct access from the voice recognition unit (i.e., not only within modality, but also across modality).

As we used a relatively limited number of voices (24 in Experiment 1 and 20 in Experiment 3), caution is required when generalizing from our findings obtained with this small set to the general interplay between voices and faces. With this limitation in mind, we argue that our results provide evidence for crossmodal interactions between unfamiliar voices and faces and, more specifically, that the properties of the voice can influence the recognition of related faces. These results extend our previous findings (Bülthoff & Newell, 2015) by providing evidence from a priming paradigm (Tulving & Schacter, 1990) that interactions between recently learned voices and faces occur at the level of implicit memory (Schacter, 1992). Finally, our findings suggest that interactions between voices and faces occur early on in information processing, possibly mediated between brain regions subserving voice and face perception (von Kriegstein, Giraud, et al., 2006), and have implications for current models of person perception.

Acknowledgments

We thank Daniel Berger for help creating the auditory stimuli, and Franziska Hausmann, Karin Bierig and Saskia Kühnhold for help in programming the experiments and testing participants. We are grateful to our colleagues for allowing us to record their voices for use as stimuli.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Max-Planck-Gesellschaft and by funding from Science Foundation Ireland [Grant No. 10/IN.1/I3003] awarded to FNN.

References

- Barenholtz, E., Lewkowicz, D. J., Davidson, M., & Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, 21(5), 1346–1352. doi:10.3758/s13423-014-0612-7
- Barsics, C., & Brédart, S. (2012). Recalling semantic information about newly learned faces and voices. *Memory*, 20(5), 527–534.
- Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20(5), 585–593.
- Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology*, 21(7), 1013–1021.
- Blank, H., Anwender, A., & von Kriegstein, K. (2011). Direct structural connections between voice-and face-recognition areas. *The Journal of Neuroscience*, 31(36), 12906–12915.
- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on computer graphics and interactive techniques – SIGGRAPH '99* (pp. 187–194). New York, NY: ACM Press. doi:10.1145/311535.311556
- Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A., & Ellis, A. W. (1991). Understanding covert recognition. *Cognition*, 39(2), 129–166.
- Bruce, V., Burton, A. M., Carson, D., Hanna, E., & Mason, O. (1994). Repetition priming of face recognition. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 179–201). Cambridge, MA: MIT Press.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305–327.
- Bülthoff, I., & Newell, F. N. (2015). Distinctive voices enhance the visual recognition of unfamiliar faces. *Cognition*, 137C, 9–21. doi:10.1016/j.cognition.2014.12.006
- Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23(1), 1–31.
- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, 81(3), 361–380. doi:10.1111/j.2044-8295.1990.tb02367
- Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. I. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology*, 87, 141–162. doi:10.1111/j.2044-8295.1996.tb02581.x
- Calder, A. J., Rhodes, G., Johnson, M. H., Haxby, J. V. (Eds.). (2011). *The Oxford handbook of face perception*. Oxford: University Press.
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8), 641–651.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11(12), 535–543.
- Casey, S. J., & Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*, 45(3), 506–513.
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, 4(1), 135–150. doi:10.1111/j.1756-8765.2011.01177.x

- Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173.
- 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 Section A*, *42*(3), 495–512. doi:10.1080/14640749008401234
- Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 193–210. doi:10.1080/14640748708401784
- Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology*, *88*(1), 143–156. doi:10.1111/j.2044-8295.1997.tb02625.x
- Graf, A. B. A., & Wichmann, F. A. (2002). Gender classification of human faces. *Biologically Motivated Computer Vision 2002 LNCS*, 2525, 491–501. doi:10.1007/3-540-36181-2_49
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337.
- Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology*, *53A*, 1105–1116.
- Hanley, J. R. (2014). Tip of the tongues for proper names. In Schwartz B., & Brown A. S. (Eds.), *Tip of the tongue and related phenomena* (pp. 50–74). Cambridge: Cambridge University Press.
- 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., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*(1), 59–67.
- Joassin, F., Pesenti, M., Maurage, P., Verreckt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex*, *47*(3), 367–376.
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596.
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-bateson, E. (2003). “Putting the face to the voice”: Matching identity across modality. *Current Biology*, *13*(19), 1709–1714. doi:10.1016/j.cub.2003.09.005
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.
- Kilgour, A. R., & Lederman, S. J. (2002). Face recognition by hand. *Perception & Psychophysics*, *64*(3), 339–352.
- Manly, T., Heutink, J., Davison, B., Gaynord, B., Greenfield, E., Parr, A., ... Robertson, I. H. (2004). An electronic knot in the handkerchief: “Content free cueing” and the maintenance of attentive control. *Neuropsychological Rehabilitation*, *14*(1–2), 89–116. doi:10.1080/09602010343000110
- Mavica, L. W., & Barenholtz, E. (2012). Matching voice and face identity from static images. *Journal of Vision*, *12*(2), 1023–1023. doi:10.1167/12.9.1023
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. doi:10.1037/h0031564
- Naci, L., Taylor, K. I., Cusack, R., & Tyler, L. K. (2012). Are the senses enough for sense? Early high-level feedback shapes our comprehension of multisensory objects. *Frontiers in Integrative Neuroscience*, *6*, 82.
- Natu, V., & O’Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology*, *102*, 726–747. doi:10.1111/j.2044-8295.2011.02053.x
- Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., ... Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *223*(1232), 293–317.
- Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493.
- Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces?: A PET study of face categorical perception. *Journal of Cognitive Neuroscience*, *13*, 1019–1034. doi:10.1162/089892901753165917
- Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist*, *47*(4), 559–569.
- Schweinberger, S. R., Herholz, A., & Stief, V. (1997). Auditory long term memory: Repetition priming of voice recognition. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 498–517.
- Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology*, *60*(10), 1446–1456. doi:10.1080/17470210601063589
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain*, *115*(1), 15–36.
- Smith, H. M. J., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016a). Concordant cues in faces and voices: Testing the backup signal hypothesis. *Evolutionary Psychology*, *14*(1), 1–10. doi:10.1177/1474704916630317
- Smith, H. M. J., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016b). Matching novel face and voice identity using static and dynamic facial images. *Attention, Perception & Psychophysics*, *78*(3), 868–879. doi:10.3758/s13414-015-1045-8
- Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. *British Journal of Psychology*, *105*(1), 1–16.
- Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face

- and voice recognition. *Psychological Research*, 77(2), 167–175.
- ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2016). The COGs (context, object, and goals) in multisensory processing. *Experimental Brain Research*, 234(5), 1307–1323.
- Troje, N. F., & Bühlhoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, 36(12), 1761–1771. doi:10.1016/0042-6989(95)00230-8
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940), 301–306.
- von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A. L., Kell, C. A., Grüter, T., Kleinschmidt, A., & Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences*, 105(18), 6747–6752.
- von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, 4(10), e326. doi:10.1371/journal.pbio.0040326
- von Kriegstein, K., Kleinschmidt, A., & Giraud, A. L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex*, 16(9), 1314–1322.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17(3), 367–376.
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, 102(4), 959–974.