Putting the pieces together: Revealing face-voice integration through the

facial overshadowing effect.

Rebecca J Tomlin, Sarah V Stevenage\* and Sarah Hammond

Department of Psychology, University of Southampton, UK

Word Count: 9,596 including abstract, tables and references

Acknowledgement of Funding:

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/J004995/1 awarded to the second author.

Correspondence:

\*Please address all correspondence to Sarah Stevenage at:

Department of Psychology, University of Southampton, Highfield, Southampton, Hampshire, SO17 1BJ

Email:svs1@soton.ac.uk, Fax: 02380 594597, Tel: 02380 592973

Abstract

Two experiments are reported which explore the limits of the facial overshadowing effect on voice recognition. In Experiment 1, the facial overshadowing effect was replicated. Voice recognition was significantly impaired when the voice had been studied in the presence of the face rather than when studied alone. However, this effect was removed when the accompanying face was inverted. In Experiment 2, the facial overshadowing effect was replicated using more realistic moving faces rather than static faces through presentation of a video sequence at study. Surprisingly, the magnitude of the facial overshadowing effect was not influenced by whether faces were static or dynamic. Moreover, the effect remained even under temporal asynchrony. Together, these results show the facial overshadowing effect to be a robust phenomenon. However, it appears to depend critically on one factor: the *facial* nature of the distraction. The results are discussed in terms of the integration of faces and voices in a recognition context, and in terms of a holistic view of face-voice processing in a more general person perception framework. This offers an explanation of current findings, and serves to guide predictions for future work.

Keywords:

Facial Overshadowing Effect, Face-Voice integration, Holistic Person Perception.

Putting the pieces together: Revealing face-voice integration through the

facial overshadowing effect.

A considerable degree of interest has been paid in recent years to the area of voice recognition. Along with face recognition, voice recognition plays a significant role in identification and, very sensibly, voice recognition researchers have taken the initial stance of applying face recognition models and methodologies to the exploration of voice recognition skills. The consistent view from this work is that the human listener is capable of recognising an individual from their voice, however, voice recognition is substantially more difficult than face recognition (see Yarmey, 1995 for a review). As a consequence, a body of work now exists which demonstrates the voice to be a relatively weak cue to person recognition.

In a realistic context, this relative difficulty of voice recognition may be offset by the relative strength of face recognition because, of course, the voice and the face are so often encountered together. In addition, given the facilitation effects often associated with multi-model over unimodal presentations (see Campanella & Belin, 2007), it is plausible to ask whether voice recognition may be improved when the voice is presented alongside the face rather than when presented alone (see Zäske, Mühl, & Schweinberger, 2015). Curiously however, the literature is mixed on the question of whether co-presentation helps or hinders subsequent voice recognition. The purpose of the present paper is to examine the evidence on this point, and to provide a framework for the results that follow.

*The Facial Overshadowing Effect*

Early work suggested that voice recognition may be facilitated by the co-presentation of the face at study. For instance, Sheffert and Olson (2004) showed that when trying to learn voices, performance was improved following study of the face and voice together (audio-visual study) compared to the study of the voice alone. However, Legge, Grossman, and Pieper (1984) and Armstrong and McKelvie (1996) both indicated no such improvement when voice recognition was tested following study of the audio-visual combination compared to study of the voice alone.

The presence of floor effects may have affected performance in previous studies. When these issues were addressed, the results of Cook and Wilding (1997) provided the first clear demonstration of a significant impairment in voice recognition when the face and voice were studied together rather than when the voice was studied alone. This has become known as the facial overshadowing effect. In this key study, participants completed a line-up task one week after presentation of a target. Study and test conditions varied in a complex design, but included (i) study via audio-visual presentation followed by a voice test; and (ii) study via audio-only presentation followed by a voice test. The results indicated a measurable drop in performance when the face accompanied the voice at study.

A series of studies now exist which also show an overshadowing effect. Notably, McAllister, Dale, Bregman, McCabe, and Cotton (1993) also used a line-up task, however, an elegant design enabled them to determine whether audio-visual input was as detrimental to subsequent *face recognition* as it was to *voice recognition*. In Experiment 1, participants completed a face recognition task whilst in Experiment 2, participants completed a voice recognition task. The results were clear in demonstrating an impairment for both recognition tasks following audio-visual rather than single-channel input. However, the impairment was greater for voice recognition than for face recognition suggesting that voice recognition was more susceptible to interference from the face than vice versa. This was replicated by Stevenage, Howland, and Tippelt (2011) using an old/new recognition task at test, providing rigour and convergent evidence from a different methodology.

 Accounting for these results, Cook and Wilding (1997) and Stevenage et al. (2011) suggested an unstoppable distraction within an audio-visual presentation such that attention was directed away from the voice and towards the attention-grabbing face. Consequently, voice recognition was impaired as it essentially became a secondary task. Cook and Wilding (2001) provided support for this explanation. In the first of two experiments, they explicitly instructed participants to attend to the voice within the audio-visual presentation as a way to offset attentional capture by the face. Despite this instruction, voice recognition performance was still impaired following audio-visual study relative to audio-only study, reinforcing the view that attention to the face was automatic and unstoppable. Interestingly, however, when Cook and Wilding’s participants were pre-exposed to the face prior to the presentation of the voice at study (Experiment 2), the facial overshadowing effect was removed. This may have reflected a degree of habituation to the face, allowing attention to be released back to the voice when subsequently presented for study, and thus supporting a better level of performance at test.

 Other methods to reduce or remove the facial overshadowing effect have also relied on manipulations that weaken attention to the face within the audio-visual presentation. Yarmey (1986), for example, reduced the illumination of the face (to simulate dusk conditions), with the expectation that participants would switch their attention to the voice as the more discernible stimulus. Similarly, Heath and Moore (2011) used faces that were partially occluded by a balaclava. Despite these manipulations, a facial overshadowing effect still emerged, with participants potentially working particularly hard to make sense of the sub-optimal facial images in each case.

 Taken together, these results suggest the facial overshadowing effect to be a robust phenomenon in which co-presentation of the face and voice at study weakens subsequent voice recognition compared to the presentation of the voice alone. The dominant explanation relies on the automatic and unstoppable attention to the face over the voice during audio-visual presentation, and this explanation has a great deal of intuitive appeal. However, there is value in ruling out an alternative and simpler explanation. Namely, as predominantly visual beings, attention may be captured by any complex visual stimulus that accompanies the voice. In this sense, the facial overshadowing effect may actually merely be a *visual* overshadowing effect. The purpose of Experiment 1 is to test this possibility.

 Experiment 1 used a sequential same/different matching task as a way of testing voice recognition following either audio-visual or audio-only study. During audio-visual study, the face was either presented in the standard upright orientation, or in an inverted orientation. Thus, across conditions, the level of visual complexity was held constant, but the perception of face-ness was reduced (see Yin, 1969). If the overshadowing effect is due to the attentional capture of the face, then it should emerge only when the face is upright. If, however, the overshadowing effect is due to the visual complexity of the accompanying stimulus, then it should emerge whether the face is upright or inverted.

*Experiment 1: Method*

*Design*

Participants took part in a sequential same/different matching task in which facial presence (present, absent) and trial type (same, different) were varied within participants, and face orientation (upright, inverted) was varied between participants. Accuracy of response on ‘same’ and ‘different’ trials represented the dependent variable, and a facial overshadowing effect would be demonstrated if performance was better when the face was absent rather than present at study.

*Participants*

A total of 64 participants (24 males, 40 females) took part on a volunteer basis or in return for course credit. Half of the participants were randomly allocated to see upright faces at study (12 males, 20 females; Mean Age = 21.75 years, *SD* = 1.95). The remaining participants were presented with inverted faces at study (12 males, 20 females; Mean Age = 22.08 years, *SD* = 1.53). All participants were unfamiliar with the stimuli used, and all had normal, or corrected-to-normal, hearing and vision.

*Materials*

The faces and voices of 32 individuals (16 males, 16 females) were used as targets in the current study. In addition, the voices of 16 individuals (8 males, 8 females) were used as foils in the ‘different’ trials. All stimuli were drawn from a bespoke database of 117 individuals collected as part of the SuperIdentity Project (www.southampton.ac.uk/superidentity) and all were young Caucasians speaking with an unremarkable southern British accent. These stimuli were selected on the basis of their ratings on a 7 point scale of vocal distinctiveness (where 1 = typical and 7 = distinctive), as provided by two independent judges. According to these ratings, the 32 targets were identified as typical-sounding (*M* = 2.46, *SD* = .65) with no voice receiving a rating greater than 3.5. Thus, any potential effects due to vocal distinctiveness were minimised. The 16 foils were identified as neither typical nor distinctive (*M* = 3.97, *SD* = .32).

 *Faces*: For each of the 32 target individuals, one static colour image of the face was obtained and this was presented alongside the voice in the study phase of face-present trials. Photographs depicted the individual in a neutral, full-face pose against a plain white background, and all images measured 6cm x 4cm. Two instances of these photographs were generated, with one depicting the face in an upright orientation, and the other being rotated through 180° to depict the face in an inverted orientation.

*Voices*: For each of the 32 target individuals, four voice clips were also obtained. The use of four clips within the current experiment ensured that different clips could be presented at study and test phases of both face-present and face-absent trials. This ensured that simple perceptual matching could not drive performance in the voice recognition task. In contrast, two clips were obtained for each of the 16 foil speakers in which they uttered the same test phrases as the targets. This enabled their presentation at the test stage of the ‘different’ trials. All clips took 4-5 seconds to utter, and were processed within Audacity 2.1.0 to ensure that any extended silences were removed, and volume was equivalent across clips.

*Phrases:* The four clips involved the targets uttering set phrases drawn from a larger sample of phrases used in the FLR2011 database (Centre for Applied Science and Technology, Home Office, UK), all of which were designed to be rich in the phonetic cues elicited from the speaker. The study phrases were ‘The smell of freshly ground coffee never fails to entice me into the shop’/ ‘The length of her skirt caused the passers-by to stare’, and the test phrases were ‘The most important thing to remember is to keep calm and stay safe’/ ‘They launched into battle with all the forces they could muster’.

These stimuli were used to create a total of 64 same/different matching trials through the presentation of all 32 targets in both face-present and face-absent conditions. Within both conditions, there were 16 ‘same’ trials and 16 ‘different’ trials. ‘Same’ trials were created by pairing the study clip of a target speaker with the test clip of the same target speaker. In contrast, ‘different’ trials were created by pairing the study clip of a target speaker with the test clip of a same-sex foil speaker. Care was taken to balance the number of male and female speakers across ‘same’ and ‘different’ trials, and to minimise item effects by counterbalancing the identity of speakers in each trial type across the participant group. Finally, for half the participants the faces were shown upright during the face-present study phase, whilst for the remaining participants, the faces were shown inverted.

 Stimuli were presented, and data were recorded via SuperLab 4.5, on an HP laptop running Windows 8.1. A 15” colour monitor with a screen resolution of 1366 x 768 pixels displayed all visual stimuli and instructions, and out-of-ear noise-cancelling headphones ensured that voice clips could be heard with a minimum of acoustic distraction.

*Procedure*

Participants were tested individually following an explanation of the study, and the provision of informed consent. The study itself consisted of a practice phase during which participants were first encouraged to orient to the response keys, and were then introduced to the auditory task. The response key orientation involved participants completing 20 trials involving the presentation of the words ‘same’ or ‘different’ on screen. Participants were asked to press ‘S’ or ‘D’ for ‘same’ and ‘different’ trials respectively and feedback was provided for both correct and incorrect responses. The auditory task orientation involved participants completing a further 20 practice trials involving the sequential presentation of two brief audio beeps. Participants were asked to press ‘S’ if the two beeps were of the ‘same’ pitch, and ‘D’ if the two beeps were of ‘different’ pitches.

 Following this practice phase, participants were introduced to the main experiment. This consisted of two blocks of 32 trials (16 ‘same’, 16 ‘different’) with the two blocks varying only in whether the face was present or absent at the study phase. Trials were presented in a random order within each block, and order of the two blocks was counterbalanced across participants. Finally, the presentation of the two blocks was separated by a self-paced break to minimise fatigue or interference effects.

Within each block, all trials took the same format as summarised in Figure 1. Trials were initiated by a 250ms presentation of a ‘get ready’ prompt, followed by a brief gap of 100ms. Following this, the study voice was presented, and this was accompanied by the study face in face-present trials. After a silent gap of 5 seconds, the test voice was presented and participants were asked to indicate whether this voice was the ‘same’ speaker, or a ‘different’ speaker, to that presented at study by pressing either ‘S’ or ‘D’ respectively. No feedback was provided on these trials. The entire study lasted 15-20 minutes, after which participants were thanked and debriefed.

(Please insert Figure 1 about here)

Results and Discussion

Given the use of a two alternative forced choice (same/different) response and the potential for bias in responding, the signal detection framework of Green and Swets (1966) was used to generate overall measures of sensitivity of discrimination (d’) and bias (C). By-participants, and by-items analyses were conducted on these variables as primary measures of performance, with accuracy providing a secondary analysis. Preliminary analysis revealed no outliers when the data were considered by-items. However, there was one outlier when the data were considered by-participants, with particularly poor performance from this participant under the most optimal condition when the face was absent from study. The data from this one participant were dropped from all subsequent analyses without replacement. A facial overshadowing effect would be demonstrated if performance in the voice recognition test was better when the face had been absent, rather than present, at study.

*Sensitivity of Discrimination (d’):* Performance was examined first through consideration of sensitivity of discrimination (see Figure 2). First, one-sample t-tests demonstrated above-chance performance in all conditions (by-participants: *t*(30) > 17.66, *p* < .001; by-items: *t*(31) > 15.09, *p* < .001). This confirmed that participants were able to do the voice-recognition task.

A 2 x 2 Analysis of Variance (ANOVA) was conducted using face presence (present, absent) and face orientation (upright, inverted) as variables. This revealed neither a main effect of face presence (by-participants: *F*(1, 61) < 1, *p* = .55; by-items: *F*(1, 31) < 1, *p* = .37) nor a main effect of face orientation (by-participants: *F*(1, 61) < 1, *p* = .62; by-items: *F*(1, 31) = 3.77, *p* = .061). However, the predicted interaction between face presence and face orientation was significant(by-participants: *F*(1, 61) = 8.99, *p* = .004, partial η2 = .13; by-items: *F*(1, 31) = 19.96, *p* < .001; partial η2 = .39).

Explorations of the simple main effects confirmed that performance was significantly better when the face was absent from study than when present (the facial overshadowing effect), but only when the faces were upright (by-participants: *F*(1, 61) = 6.39, *p* = .014; by-items: *F*(1, 31) = 10.81, *p* = .003). This effect disappeared when faces were inverted (by-participants: *F*(1, 61) = 2.91, *p* = .093; by-items: *F*(1, 31) = 3.43, *p* = .074).

(Please insert Figure 2 about here)

*Accuracy of Performance:* Table 1 summarises the accuracy across ‘same’ and ‘different’ trials in each experimental condition. In order to determine whether the facial overshadowing effect applied to both ‘same’ and ‘different’ trials, a 2 x 2 x 2 ANOVA was conducted on the accuracy of performance, using face presence (present, absent), trial type (same, different) and face orientation (upright, inverted) as variables. As with the analysis of d’ above, this revealed neither a main effect of face presence (by-participants: *F*(1, 61) < 1, *p* = .39; by-items: *F*(1, 31) < 1, *p* = .56) nor a main effect of face orientation (by-participants: *F*(1, 61) < 1, *p* = .44; by-items: *F*(1, 31) < 1, *p* = .39). However, the predicted interaction between face presence and face orientation again emerged (by-participants: *F*(1, 61) = 9.08, *p* = .004, partial η2 = .13; by-items: *F*(1, 31) = 20.66, *p* < .001, partial η2 = .40).

Exploration of the simple main effects confirmed the presence of a facial overshadowing effect when faces were upright (by-participants: *F*(1, 61) = 7.44, *p* = .008; by-items: *F*(1, 31) = 8.85, *p* = .006) but not when faces were inverted (by-participants: *F*(1, 61) = 2.32, *p* = .133; by-items: *F*(1, 31) = 3.20, *p* = .08).

In terms of the effect of trial type, the ANOVA revealed a main effect in the by-participants analysis only (by-participants: *F*(1, 61) = 16.02, *p* < .001, partial η2 = .21; by-items: *F*(1, 31) = 2.28, *p* = .14), with performance being better on ‘same’ trials (*M* = .91) than ‘different’ trials (*M* = .87). However, trial type did not interact with any other variable (by-participants: all *F*s(1, 61) < 2.33, *p* > .13; by-items: *F*s(1, 31) < 2.11, *p* > .16), and thus did not moderate the demonstration of the facial overshadowing effect itself. In this regard, the facial overshadowing effect emerged for both ‘same’ and ‘different’ trials alike.

(Please insert Table 1 about here)

*Response Bias (C):* Finally, the data pertaining to response bias were considered. A one-sample *t*-test revealed an overall tendency to say ‘same’ rather than ‘different’ when data were considered by-participants (by-participants: *M* = -.16; *t*(62) = 4.41, *p* < .001; by-items: *M* = -.08; *t*(31) = 1.19, *p* = .24). In order to determine whether the facial overshadowing effect reflected a bias in responding, a 2 x 2 ANOVA was conducted on bias (C) using face presence (present, absent) and face orientation (upright, inverted) as variables. This revealed neither a main effect of face presence (by-participants: *F*(1, 61) < 1, *p* = .55; by-items: *F*(1, 31) = 3.98, *p* = .055), nor a main effect of face orientation (*F*(1, 61) < 1, *p* = .55; by-items: *F*(1, 31) = 1.92, *p* = .18). Additionally, there was no interaction between face presence and orientation (by-participants: *F*(1, 61) = 2.96, *p* = .09; by-items: *F*(1, 31) = 1.92, *p* = .18). Consequently, the facial overshadowing effect did not appear to be due to a shift in the response criterion when the face was present or absent. Instead, it appeared to be attributable to perceptual factors associated with the task itself.

 Taken together, the results of Experiment 1 confirmed the presence of a facial overshadowing effect, with poorer performance on a voice recognition task when the voice had been presented alongside a face at study compared to when presented alone. Notably, however, this facial overshadowing effect disappeared when the face was inverted, and this was shown when the data were considered by-participants and by-items. Given that the inverted face retained all of the visual complexity of the upright face, but was harder to process ‘as a face’, these results may be taken to indicate that the overshadowing effect is dependent on *facial* rather than mere *visual* distraction.

Experiment 2

The results of Experiment 1 showed the facial overshadowing effect to be robust and replicable, but also showed it to be dependent on distraction at study by a *face* rather than by any complex visual stimulus. The literature to date has suggested that this facial overshadowing effect may occur because the face is data-rich (Cook & Wilding, 1997; Stevenage et al., 2011), and thus grabs attention from the voice at study, leaving voice recognition impaired at test.

The purpose of Experiment 2 was to explore the facial overshadowing effect further. In particular, it was hypothesized that the extent of the overshadowing effect may be modified if the accompanying face was dynamic rather than static. This assumption rests on demonstration that a moving face may provide a richer source of information than a static face and may thus have a greater disruptive influence on voice processing (Lander & Chuang, 2005). In addition, it was hypothesized that the facial overshadowing effect may be maximised if conditions permitted strong audio-visual integration. Integration at the identity level may occur through co-presentation of the face and voice whereas integration at the articulatory level relies on this co-presentation being in synchrony, or within a small temporal window, rather than being temporally offset (see Robertson & Schweinberger, 2010).

Given these two questions, Experiment 2 explored the facial overshadowing effect using dynamic rather than static facial images and manipulated the presentations such that the face and voice were either in synchrony or not. Comparison of the dynamic synchronous condition with the static (synchronous) condition described in Experiment 1 allowed a test of the first question. Comparison of the dynamic synchronous condition with the dynamic asynchronous condition allowed a test of the second question.

Method

*Design*

A sequential same/different matching task was again used to examine the facial overshadowing effect. As in Experiment 1, a 2 x 2 x 2 mixed design was used in which facial presence (present, absent), and trial type (same, different), were varied within participants. The between-participants variable was temporal synchrony (synchronous, asynchronous). Accuracy of response on ‘same’ and ‘different’ trials represented the dependent variable, and a facial overshadowing effect would be demonstrated if performance was better when the face was absent rather than present at study.

*Participants*

A total of 72 participants (14 males, 58 females) took part on a volunteer basis or in return for course credit. Half of the participants were randomly allocated to see moving faces in synchrony with the voices at study (6 males, 30 females; Mean Age = 19.6 years, *SD* = 1.66). The remaining participants were presented with moving faces which were asynchronous with the voices at study (8 males, 28 females; Mean Age = 20.8 years, *SD* = 5.53). All participants had normal, or corrected-to-normal, hearing and vision and none were familiar with the stimuli used. Additionally, none had taken part in Experiment 1.

*Materials*

 The stimuli were identical to those used in Experiment 1 with the exception that video sequences were used rather than static photographs during the study phase of the face-present trials. These video sequences were recorded with a Panasonic AG-HMC41 HD video camera positioned at a distance of 3m from the target speaker, with a directional microphone attached to the camera, and a lapel microphone positioned on the speaker. The video sequences depicted the speaker reading the set phrases from a teleprompter. The audio clips for the corresponding face-absent trials were identical to those used in Experiment 1, having been extracted from these video sequences.

Synchronous face-voice presentations were achieved merely by playing the video clip with the sound audible. However, in order to create the asynchronous face-voice presentations, the voice clip was initiated 1000ms in advance of the silent video clip with the result that the face and voice were co-presented (but offset) for 3-4 seconds rather than for the full 4-5 second clip[[1]](#footnote-1). The use of a 1000ms asynchrony here was selected to fall outside the audio-visual integration window indicated by Robertson and Schweinberger (2010). Nevertheless, the length of the stimuli meant that even with this asynchrony, the face and voice overlapped by 3-4 seconds potentially allowing some level of audio-visual integration. However, this level of audio-visual integration was anticipated to be below that possible when face and voice were presented in synchrony.

Stimulus presentation and data recording were controlled using SuperLab 4.5 running on a DELL laptop with a 15” colour monitor and a screen resolution set to 1366 x 768 pixels. As in Experiment 1, audio clips were presented via out-of-ear noise-cancelling headphones to ensure that voice clips could be heard with a minimum of acoustic distraction.

*Procedure*

 All participants were tested individually using a procedure which mirrored that used in Experiment 1. In this regard, participants were briefed on the nature of the task prior to providing informed consent. They then engaged in a practice phase which was identical to that in Experiment 1, before undertaking the main phase of the experiment.

 The main phase took the same format as Experiment 1 and consisted of two blocks of 32 trials (16 ‘same’ trials, 16 ‘different’ trials) separated by a self-paced break. The two blocks of trials differed only in terms of whether the video was presented at study (face-present) or not (face-absent), and order of these two blocks was counterbalanced across participants.

 As before, all trials within the blocks took the same format as illustrated in Figure 3, consisting of the initial presentation of a ‘get ready’ prompt for 250ms, followed by a brief gap of 100ms. This was followed by the study clip presented either as an audio-only clip or as a video clip. After a 5 second gap, the test voice was presented, and participants were asked to indicate whether this voice was the ‘same’ speaker, or a ‘different’ speaker, to that presented at study by pressing either ‘S’ or ‘D’ respectively. No feedback was provided, and the study lasted 15-20 minutes after which participants were thanked and debriefed.

(Please insert Figure 3 about here)

Results and Discussion

As in Experiment 1, the facial overshadowing effect was explored through analysis of the primary measures of sensitivity of discrimination (d’) and response bias (C), with analysis of accuracy presented as a secondary analysis. Again, analyses were conducted both by-participants and by-items. Preliminary exploration of the data revealed 3 outlier participants (1 in the synchronous condition; 2 in the asynchronous condition). In addition, there were 2 outlier items. All outliers reflected particularly poor performance under the most optimal condition when the face was absent from study. The data from all outliers were removed from all analyses without replacement.

The current data enabled the examination of two questions pertaining to whether facial overshadowing was increased with dynamic rather than static faces; and whether facial overshadowing was increased with synchronous rather than asynchronous dynamic presentations. For both questions, a facial overshadowing effect would be demonstrated if performance in the voice recognition test was better when the face had been absent, rather than present, from study.

*Dynamic versus Static Faces.*

 *Sensitivity of Discrimination:* As in Experiment 1, performance was examined first through consideration of sensitivity of discrimination (see Figure 4). One sample t-tests confirmed above-chance performance in all conditions (by-participants: *t*s > 17.66, *p* < .001; by-items: *t*s > 18.36, *p* < .001). This confirmed that participants were again able to do the voice-recognition task across all conditions.

A 2 x 2 ANOVA was conducted using face presence (present, absent) and facial motion (static, dynamic) as variables. This revealed a main effect of face presence (by-participants: *F*(1, 63) = 14.70, *p* < .001, partial η2 = .189; by-items: *F*(1, 29) = 8.85, *p* = .006, partial η2 = .234), with performance being better when the face was absent rather than present from study (the facial overshadowing effect). There was also a main effect of facial motion (by-participants: *F*(1, 63) = 19.91, *p* < .001, partial η2 = .240; by-items: *F*(1, 29) = 12.59, *p* = .001, partial η2 = .303) which amounted to better overall performance by those taking part in Experiment 2 than Experiment 1. Importantly, however, the demonstration of the facial overshadowing effect was not affected by the motion (or otherwise) of the face (by-participants: *F*(1, 63) < 1, *p* = .62; by-items: *F*(1, 29) = 1.49, *p* = .23) indicating that the facial overshadowing effect was no stronger when the accompanying face was dynamic rather than static.

(Please insert Figure 4 about here)

 *Accuracy:* Tables 1 and 2 summarise the accuracy across ‘same’ and ‘different’ trials when faces were either static or dynamic and when the face was either present at study or not. A 2 x 2 x 2 ANOVA was used to explore whether the facial overshadowing effect demonstrated above applied to both ‘same’ and ‘different’ trials. As such, face presence (present, absent), facial motion (static, dynamic) and trial type (‘same’, ‘different’) were used as variables. The analysis revealed a main effect of face presence (by-participants: *F*(1, 63) = 12.20, *p* = .001, partial η2 = .162; by-items: *F*(1, 29) = 6.41, *p* = .017, partial η2 = .181) with greater accuracy when the face was absent than when present (the facial overshadowing effect). There was also a main effect of facial motion (by-participants: *F*(1, 63) = 17.38, *p* < .001, partial η2 = .216; by-items: *F*(1, 29) = 7.52, *p* = .01, partial η2 = .206) amounting to greater accuracy overall by those taking part in Experiment 2 than Experiment 1. However, these two factors did not interact (by-participants: *F*(1, 63) < 1, *p* = .72; by-items: *F*(1, 29) < 1, p *=* .73) indicating that the facial overshadowing effect was again not affected by the motion (or otherwise) of the face.

In terms of trial type, the analysis revealed no main effect of trial type overall (by-participants: *F*(1, 63) = 1.44, *p* = .24; by-items: *F*(1, 29) < 1, *p* = .74). However, trial type did interact with facial motion (by-participants: *F*(1, 63) = 10.32, = *p* = .002, partial η2 = .141; by-items: *F*(1, 29) = 9.42, *p* = .005, partial η2 = .245) amounting to better performance on ‘same’ than ‘different’ trials in Experiment 1 but the reverse pattern in Experiment 2. Importantly, however, trial type did not interact with face presence (by-participants: *F*(1, 63) < 1, *p*= .79; by-items: *F*(1, 29) < 1, *p* = .77), and the three-way interaction between all variables approached but did not reach significance across both analyses (by-participants: F(1, 63) = 4.00, *p* = .05; by-items: *F*(1, 29) = 3.19, *p* = .08). Given the opportunity for a three-way interaction to represent noise in the data, the conservative interpretation is that the facial overshadowing effect was not modified by trial type, facial motion, or their combination.

 *Response Bias:* Finally, the pattern of response bias was considered. A one-sample *t*-test revealed some evidence of a tendency to say ‘same’ rather than ‘different’ in Experiment 1 (by-participants: *t*(30) = 3.3, *p* = .001; by-items: *t*(29) = 1.5, *p* = .14) but no bias in Experiment 2 (by-participants: *t*(33) = 1.36, *p* = .19; by-items: *t*(29) = 1.79, *p* = .08). A 2 x 2 ANOVA explored the effect of response bias on the facial overshadowing effect, using face presence (present, absent) and facial motion (static, dynamic) as variables. This revealed no effect of face presence (by-participants: *F*(1, 63) < 1, *p* = .83; by-items: *F*(1, 29) < 1, *p* = .39). However, there was a main effect of facial motion (by-participants: *F*(1, 63) = 12.11, *p* = .001, partial η2 = .161; by-items: *F*(1, 29) = 13.84, *p* = .001, partial η2 = .323) amounting to a difference across Experiments 1 and 2. Curiously, an interaction between face presence and facial motion also emerged (by-participants: *F*(1, 63) = 6.48, *p* = .013, partial η2 = .093; by-items: *F*(1, 29) = 4.85, *p* = .04; partial η2 = .143). Examination of the simple main effects suggested that this reflected a difference in response bias across the two Experiments when the face was absent despite the fact that these effectively represented identical conditions (by-participants: *t*(63) = 4.47, *p* < .001; by-items: *t*(29) = 4.19, *p* < .001 ). In contrast, there was no difference in response bias across the two Experiments when the face was present (by-participants: *t*(63) = 1.09, *p* = .28; by-items: *t*(29) = 1.08, *p* = .29). These data are difficult to interpret. However, given that the difference in response bias across facial motion emerges in the condition in which faces are not actually presented, it most likely reflects variation across the two groups of participants rather than anything more meaningful.

In summarising the results across static and dynamic faces, it had been anticipated that the facial overshadowing effect may be magnified when the accompanying faces at study were dynamic rather than static given the strong audio-visual integration that becomes possible in the former case. Surprisingly, the results did not bear out this expectation. Instead, the facial overshadowing effect emerged with equivalent magnitude regardless of whether the accompanying faces were static or dynamic. At this stage, these results suggested that the facial overshadowing effect may rely solely on co-presentation of face and voice. Analysis of the second question relating to synchronous and asynchronous dynamic presentations will be important in verifying this interpretation.

*Synchronous and Asynchronous Dynamic Presentations*

*Sensitivity of Discrimination:* As before, performance was examined first through consideration of sensitivity of discrimination (see Figure 5). One-sample t-tests demonstrated above-chance performance in all conditions (by-participants: *t*(33) > 24.46, *p* < .001; by-items: *t*(29) > 19.28, *p* < .001) demonstrating that participants were again able to do the voice-recognition task across all conditions.

A 2 x 2 mixed ANOVA was conducted using face presence (present, absent) and face-voice synchrony (synchronous, asynchronous) as variables. This revealed a significant main effect of face presence (by-participants: *F*(1, 67) = 17.52, *p* < .001, partial η2 = .207; by-items: *F*(1, 29) = 5.51, *p* = .026, partial η2 = .160) with performance being better when the face was absent rather than present from study overall (the facial overshadowing effect). There was no main effect of face-voice synchrony (by-participants: *F*(1, 67) < 1, *p* = .48; by-items: *F*(1, 29) < 1, *p* = .35). Moreover, there was no interaction between face presence and face-voice synchrony (by-participants: *F*(1, 67) < 1, *p* = .70; by-items: *F*(1, 29) < 1, *p* = .33). These results suggested that the facial overshadowing effect emerged regardless of synchronous or asynchronous presentation.

(Please insert Figure 5 about here)

*Accuracy*: As in Experiment 1, an analysis of accuracy was conducted in order to determine whether the facial overshadowing effect applied to both ‘same’ and ‘different’ trials. A 2 x 2 x 2 ANOVA explored the effects of face presence (presence, absent), trial type (same, different), and face-voice synchrony (synchronous, asynchronous) as variables. This revealed a main effect of face presence (by-participants: *F*(1, 67) = 13.96, *p* < .001, partial η2 = .172; by-items: *F*(1, 29) = 4.36, *p* = .046, partial η2 = .131) with performance being better when the face was absent from study than when present (the facial overshadowing effect). There was no main effect of face-voice synchrony (by-participants: *F*(1, 67) < 1, *p* = .83; by-items: *F*(1, 29) < 1, *p* = .90). Moreover, there was again no interaction between face presence and face-voice synchrony (by-participants: *F*(1, 67) < 1, *p* = .95; by-items: *F*(1, 29) < 1, *p* = .97). This confirmed the pattern above and suggested that the facial overshadowing effect emerged regardless of synchronous or asynchronous presentation.

 In terms of the effect of trial type, the analyses revealed no main effect of trial type on accuracy of performance (by-participants: *F*(1, 67) = 1.75, *p* = .19; by-items: *F*(1, 29) < 1, *p* = .62). Moreover, as in Experiment 1, trial type did not interact with any other variable (by-participants: all *F*s (1, 67) < 3.43, *p* > .07; by-items: all *F*s(1, 29) < 1, *p* > .37). Thus, trial type did not moderate the demonstration of the facial overshadowing effect itself.

*Response Bias (C)*: Finally, the data pertaining to response bias were once again considered. A one-sample *t*-test revealed no bias overall (by-participants: *M* = .03; *t*(68) < 1, *p* = .42; by-items: *M* = .13; *t*(29) = 1.67, *p* = .11). In order to determine whether the facial overshadowing effect here was affected by a bias in responding, a 2 x 2 ANOVA was conducted on bias (C) using face presence (present, absent) and face-voice synchrony (synchronous, asynchronous) as variables. This revealed a main effect of face presence when data were considered by-participants, with more of a bias to say ‘different’ when the face was absent rather than present at study, (by-participants: *F*(1, 67) = 5.64, *p* = .02, partial η2 = .078; by-items: *F*(1, 29) < 1, *p* = .46). There was, however, no effect of face-voice synchrony (by-participants: *F*(1, 67) = 1.34, *p* = .25; by-items: *F*(1, 29) < 1, *p* = .56). Additionally, there was no interaction between face presence and face-voice synchrony (by-participants: *F*(1, 67) < 1, *p =* .73; by-items: *F*(1, 29) < 1, *p* = .78). Consequently, there was some evidence of a shift in response bias contributing to the facial overshadowing effect but this did not emerge as a consistent finding across analyses.

 In summarising the results across synchronous and asynchronous dynamic faces, it had been anticipated that the facial overshadowing effect may be maximised when the opportunity existed to engage in strong audio-visual integration between face and voice through their synchronous presentation. As such, a stronger facial overshadowing effect was anticipated in the synchronous condition than in the asynchronous condition. Surprisingly again, the results revealed a facial overshadowing effect for both conditions regardless of the synchrony (or otherwise) of the stimuli, and the magnitude of this effect was not moderated by the timing manipulation. This suggested that the overshadowing effect did not depend on audio-visual synchrony at the articulatory level but instead on a basic co-occurrence of the face and voice.

One way of interpreting the results across synchronous and asynchronous conditions here is that the facial overshadowing effect under synchronous conditions represents the influence of the co-presentation of the face, as in Experiment 1, whilst the facial overshadowing effect under asynchronous conditions represents something else. One possibility is that the very asynchrony of the presentation created an experience that was surprising, disturbing, or hard to ignore, rather like the perception of a poorly dubbed film when speech and visual movement are offset. If this were the case, one may have anticipated a greater facial overshadowing effect in the asynchronous case (where both faceness and participant reaction were factors) than in the synchronous case. The data did not indicate any difference in magnitude of effect[[2]](#footnote-2).

A simpler explanation for the body of data provided here is that an overshadowing effect may simply emerge whenever a face accompanies the voice at study. This may arise because the co-presented face is distracting. Equally, it may arise because of a capacity to associate or bind the face and the voice together within multimodal person perception, in which the face overshadows the voice as a stronger cue to identity. Whilst the present data do not allow these mechanisms to be differentiated, the concept of integration is gaining traction within the literature. Importantly, however, the present data confirm that this integration need not be at an articulatory level through synchronised speech and lip movements. Instead, it may be at the identity level through temporal overlap of face and voice. Differentiation of the two levels of integration is discussed by Robertson and Schweinberger (2010), who suggested that audio-visual integration for the purposes of speech perception may rely on a much tighter temporal synchrony, and may tolerate a smaller departure from synchrony, compared to the audio-visual integration for the purposes of identity processing. On the basis of the present data, it is suggested that co-presentation of faces and voices, no matter if the face is static, dynamic, synchronised, or not, may enable integration for the purposes of identity processing, but carries a consequence as seen through facial overshadowing. These data sit well within a context in which face and voice processing are viewed as representing different channels in a multi-modal person perception framework. It is to this that the discussion now turns.

General Discussion

The results of two experiments have demonstrated the facial overshadowing effect to be a robust phenomenon in that voice recognition was impaired when the voice was studied alongside the face, rather than when studied alone. However, the results here also provided clarity regarding the conditions under which the facial overshadowing effect emerged. Importantly, the overshadowing effect was demonstrated when an upright face accompanied a voice at study but not when an inverted face accompanied the voice. As such, the effect was due to facial rather than visual co-presentation. Moreover, the overshadowing effect was demonstrated regardless of whether the face was static or dynamic, and regardless of whether the face and voice were presented in synchrony or in asynchrony suggesting that it did not depend on articulatory integration as might be demanded for speech perception. The present results may represent attentional distraction by the co-presented face. However, they can equally be interpreted to reflect a powerful tendency to integrate the face and voice as belonging to the same identity when they are co-presented. The face, as a dominant cue to identity, then overshadows the voice in the current recognition task.

An integration explanation for the facial overshadowing effect carries notable implications, and there is value in exploring these here. In particular, if the co-presentation of face and voice at study provides an opportunity to bind the two inputs, then it may be expected that the subsequent recognition of *either* input alone would lead to a measurable deficit. In other words, whilst audio-visual study impairs voice recognition (the facial overshadowing effect), it may also be anticipated that audio-visual study would impair face recognition (in an analogous ‘vocal overshadowing effect’). Evidence on this point is difficult to obtain given the generally high level of performance in face recognition tasks. Nevertheless, McAllister et al*.* (1993) provided evidence which accords with this prediction. In their study, a test of face recognition was impaired following audio-visual rather than visual study, and a test of voice recognition was impaired following audio-visual rather than audio study. The magnitude of the impairment was greater for voices than for faces, suggesting that voices were the more susceptible of the two modalities. Nevertheless, an impairment in the face recognition task was noted. This vocal overshadowing effect of face recognition was not replicated by Stevenage, Howland, and Tippelt (2011), potentially because of near-ceiling levels of performance in the face recognition task regardless of study conditions. This suggests that the demonstration of the fragile vocal overshadowing effect may depend entirely on the use of a task which is challenging enough to avoid ceiling levels of face recognition performance.

*A Holistic Approach to Person Perception:*

An explanation of the current results in terms of integration of faces and voices when processing identity should, perhaps, not be a surprise. Indeed, the brain areas responsible for face, and voice, recognition have been seen to interact when both stimuli are co-presented (Joassin, Pesenti, Maurage, Verreckt, Bruyer, & Campanella, 2011; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005), and direct neural links have been revealed between face and voice processing areas (Blank, Anwander, & von Kriegstein, 2011). This supports recent thinking notably by Young and Bruce (2011) regarding a holistic approach for person perception more generally. At the person perception level, people are represented through multi-modal channels, which *de-facto* demand multi-modal processing, and integration of that processing in a multi-modal framework. As such, the concept of integration is rapidly gaining traction amongst those who seek to understand person perception as a realistic and multimodal task. This is a key feature of the multimodal modal of person perception described by Yovel and O’Toole (2016). More specifically, they describe a model in which the face, voice (and body) of an individual are integrated such that different inputs assume differential importance to a recognition task over space and over time.

Integration is also a key feature of the multimodal model described by Belin and colleagues (Belin, Fecteau, & Bédard, 2004; Belin, Bestelmeyer, Latinus, & Watson, 2011). Belin et al. emphasized the fact that the face and the voice could each contribute to decisions not just related to identity, but related to affect and speech processing also. Interestingly the body of evidence that now exists suggests that whilst both face and voice contribute to decisions in all three domains, they may not contribute equally. Indeed, the face may contribute more than the voice to identity decisions, and the voice may contribute more than the face to speech decisions (see Stevenage & Neil, 2014). This holistic framework enables consideration of processing of multimodal stimuli when face and voice send the same information. Perhaps more interestingly, the framework can offer understanding when the face and the voice send conflicting information. It is to this that the discussion now turns.

The first domain in which the face and the voice may conflict is ‘identity’, and the question that arises concerns the resolution of the percept when the face of one person is presented alongside the voice of another person. Evidence already exists on this point through the work of Schweinberger, Robertson, and Kaufmann (2007) and Stevenage, Neil, and Hamlin (2014). This suggests that when the face and voice conflict, the face dominates the percept. Indeed, voice recognition is significantly impaired when the face of another person is presented, whereas face recognition is not impaired at all when the voice of another person is presented. This confirms the strength of the face over the voice as a cue to identity, and suggests that the visual input dominates within the identity domain. Within the holistic framework, the inequality of face and voice in the identity domain is denoted through the facial and acoustic inputs to the identity decision being located relatively far apart. Moreover, it is worth noting that the brain regions assumed to oversee face and voice recognition show spatial separation as well, with face recognition being largely conducted within the right fusiform face area (FFA), and voice recognition being conducted within the right anterior temporal lobe (see Gainotti, 2013 for a review).

 The second domain in which the face and voice may conflict is ‘affect’ and this may arise if the face conveys one emotion whilst the voice conveys another. The literature is well established regarding the expression of emotion through the face, and this suggests the existence of universal facial expressions (see Ekman & Friesen, 1971) complemented by a host of micro-expressions, or fleeting blends, of our cardinal displays. The literature on vocal expression is perhaps less well-established by comparison (for an overview, see Laukka, Juslin, & Bresin, 2005; Scherer, 1986). However, it is clear that both the face and voice contribute to decisions in this domain and that their combination is hugely beneficial through providing redundancy when one or other channel may be unavailable (see Campanella & Belin, 2007).

 When in conflict, however, available evidence suggests that the face again dominates the percept and is judged as more authentic. Indeed, when the facial and vocal expressions are mismatched, the categorisation of emotion is affected more by the expression seen in the face than by that heard in the voice (Collignon, Girard, Gosselin, Roy, Saint-Amour, Lassonde, & Lepore, 2008; Jacob, Kreifelts, **Brück**, Erb, Hösl, & Wildgruber, 2012). Such situations may arise in response to social display rules, or social tools such as irony, rhetoric or humour. Similarly, a mismatch between facial and vocal affect may arise when an individual seeks to deceive. In this regard, individuals may try to control their facial features but may nevertheless leak affective cues through their voice, or vice versa. If this mismatch is discernible, it may be a useful cue in the detection of deception.

 The third domain in which the face and voice may offer conflicting inputs is ‘speech’, and this may arise if the lip movements indicate one utterance whilst the voice indicates another. Experimentally, this has been demonstrated through the well-known McGurk effect in which a contradiction between the visual and auditory inputs is resolved to give a third unique perception (McGurk & McDonald, 1976). More specifically, when the lips say ‘ga’ and the voice says ‘ba’, the observer perceives the sound ‘da’. This fusion is very interesting as it suggests that neither the face nor the voice dominates the percept in the domain of speech processing. Indeed, this demonstration underlines the prioritisation of speech perception as a task to which both visual and auditory channels actively contribute. Moreover, it confirms the close functional arrangement of the face and the voice in this domain of processing.

*Conclusion*

Taken together, the results here provide greater understanding regarding the facial overshadowing effect. Specifically, the effect is confirmed to be a *facial* rather than a visual effect. Moreover, it is suggested to be a consequence of the *integration* of face and voice when processing identity, and reflects the cost when these inputs are later processed in isolation. At a broader level, however, these results may be interpreted within the context of a larger holistic framework. This recognises the perspective articulated by Belin and colleagues in which faces and voices both contribute to decisions about identity, speech and affect. In addition to accounting for the current data, the discussion here has usefully reflected on the relative dominance of the face and voice when making judgements about identity, speech and affect. This holistic framework is thus capable of accounting for a significant breadth of both neuropsychological and empirical research. However, it is hoped that it may also guide predictions for future research within a context in which person perception is recognised as a multimodal task involving the processing of a rich and integrated set of cues rather than a series of isolated ‘parts’.

Acknowledgements

The authors would like to thank Professor Andy Young and Dr Greg Neil for valuable conversations on early versions of the ideas contained here.

References

Armstrong, H.A., & McKelvie, S.J. (1996). Effect of face context on recognition memory for voices. *The Journal of General Psychology, 123(3)*, 259-270. doi: 10.1080/00221309.1996.9921278

Belin, P., Bestelmeyer, P.E.G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology. 102*, 711-725. doi: 10.1111/j.2044-8295.2011.02041.x

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Science. 8 (3*), 129-135.

Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *The Journal of Neuroscience, 7*, 12906-12915.

Campanella S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11(12),* 535-543*.* doi: 10.1016/j.tics.2007.10.001

Collingnon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-Visual Integration of Emotion Expression. *Brain Research, 1242(25)*, 126-135. DOI: http://dx.doi.org/10.1016/j.brainres.2008.04.023

Cook, S., & Wilding, J. (1997). Earwitness Testimony 2: Voices, faces and context. *Applied Cognitive Psychology, 11,* 527-541. doi: 10.1002/(SICI)1099-0720(199712)11:6<527::AID-ACP483>3.0.CO;2-B

Cook, S., & Wilding, J. (2001). Earwitness testimony: Effects of exposure and attention on the Face Overshadowing Effect. *British Journal of Psychology, 92*, 617-629.

Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* **17**: 124–129. DOI: http://dx.doi.org/[10.1037/h0030377](http://dx.doi.org/10.1037/h0030377).

FLR2011. Database created by the Centre for Applied Science and Technology, Home Office, UK.

Gainotti, G. (2013). Laterality effects in normal subjects’ recognition of familiar faces, voices and names. Perceptual and representational components. *Neuropsychologica, 51(7)*, 1151-1160. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2013.03.009

Green, D.M., & Swets J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley. ([ISBN 0-471-32420-5](https://en.wikipedia.org/wiki/Special%3ABookSources/0471324205))

Heath, A.J., & Moore, K. (2011). Earwitness memory: Effects of facial concealment on the face overshadowing effect. *International Journal of Advances Science and Technology, 33*, 131-140.

**Jacob, H., Kreifelts, B., Brück, C., Erb, M., Hösl, F., & Wildgruber, D. (2012). Cerebral integration of verbal and nonverbal emotional cues: Impact of individual nonverbal dominance. *NeuroImage, 61(3),* 738-747.**

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*, 367-376.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition, 12(3),* 429-442.

**Laukka, P., Juslin, P. & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion, 19(5)*,** 633-653, DOI: http://dx.doi.org/10.1080/02699930441000445

Legge, G.E., Grossman, C., & Pieper, C.M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning Memory and Cognition, 10*, 298-303.doi: 10.1037/0278-7393.10.2.298

McAllister, H.A., Dale, R.H.I., Bregman, N.J., McCabe, A., & Cotton, C.R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology, 14(2),* 161-170. doi: 10.1207/s15324834basp1402\_3

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

Robertson, D.M.C., & Schweinberger, S.R. (2010). The role of audio-visual asynchrony in person recognition. *The Quarterly Journal of Experimental Psychology, 63(1)*, 23-30.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143-165.

Schweinberger, S.R., Robertson, D., & Kaufmann, J.M. (2007). Hearing facial identities. *Quarterly Journal of Experimental Psychology, 60(10)*, 1446-1456. doi**:** 10.1080/17470210601063589

Sheffert, S.M., & Olson, E. (2004). Audio-visual speech facilitates voice learning. *Perception and Psychophysics, 66(2)*, 352-362.

Stevenage, S.V., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness Recognition. *Applied Cognitive Psychology, 25(1-*0, 112-118.

Stevenage, S.V., & Neil, G.J. (2014). Hearing faces and seeing voices: The integration and interaction of face and voice processing. *Psychologica Belgica, 54 (3),* 266-281. doi: <http://dx.doi.org/10.5334/pb.ar>

Stevenage, S.V., Neil, G.J., & Hamlin, I. (2014). When the face fits: Recognition of celebrities from matching and mismatching faces and voices. *Memory, 22(3)*, 284-294. doi: <http://dx.doi.org/10.1080/09658211.2013.781654>

SuperIdentity Stimulus Database: [www.southampton.ac.uk/superidentity](http://www.southampton.ac.uk/superidentity)

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. doi: 10.1162/0898929053279577

Yarmey, A.D. (1986). Verbal, visual and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology, 71(3)*, 363-370.

Yarmey, A.D. (1995). Earwitness speaker identification. *Psychology, Public Policy and Law. 1 (4),* 792-816.

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

Young, A.W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974. doi: 10.1111/j.2044-8295.2011.02045.x

Yovel, G., & O’Toole, A.J. (2016). Recognizing people in motion. *Trends in Cognitive Science, 20(5)*, 383-395. doi: 10.1016/j.tics.2016.02.005

Zäske, R., Mühl, C., & Schweinberger, S.R. (2015). Benefits for voice learning caused by concurrent faces develop over time. *PLOSOne*, November 20th, doi: 10.1371/journal-pone.0143151

Table 1: Mean response bias together with proportion accuracy for ‘same’ and ‘different’ trials in Experiment 1 following study with and without the presence of upright or inverted static faces (standard deviation is presented in parentheses).

|  |  |  |
| --- | --- | --- |
|  | Upright | Inverted |
|  | Face Present | Face Absent | Face Present | Face Absent |
| Bias (C) | -.12 (.36) | -.26 (.36) | -.18 (.40) | -.11 (.39) |
| Prop Accuracy (‘same’ trials) | .88 (.10) | .94 (.07) | .93 (.07) | .90 (.08) |
| Prop Accuracy (‘different’ trials) | .85 (.10) | .87 (.08) | .88 (.09) | .87 (.09) |

Table 2: Mean response bias together with proportion accuracy for ‘same’ and ‘different’ trials in Experiment 2 following study with and without the presence of synchronous or asynchronous moving faces (standard deviation is presented in parentheses).

|  |  |  |
| --- | --- | --- |
|  | Synchronous | Asynchronous |
|  | Face Present | Face Absent | Face Present | Face Absent |
| Bias (C) | -.01 (.43) | .16 (.38) | -.07 (.39) | .05 (.41) |
| Prop Accuracy (‘same’ trials) | .91 (.09) | .93 (.08) | .92 (.08) | .93 (.07) |
| Prop Accuracy (‘different’ trials) | .92 (.07) | .96 (.06) | .91 (.07) | .95 (.05) |

*Figure 1*: Illustration of the procedure used in Experiment 1.



*Figure 2*: Mean sensitivity of discrimination (d’) in a sequential same/different voice matching task following study with or without presentation of an upright or an inverted static face (Experiment 1). Bars represent standard deviation.

*Figure 3*: Illustration of the procedure used in Experiment 2.



*Figure 4*: Mean sensitivity of discrimination (d’) in a sequential same/different voice matching task following study with or without presentation of a static face (Experiment 1) or a dynamic face (Experiment 2). Bars represent standard deviation.

*Figure 5*: Mean sensitivity of discrimination (d’) in a sequential same/different voice matching task following study with or without presentation of a synchronous or an asynchronous moving face (Experiment 2). Bars represent standard deviation.

1. An audio lead was preferred to an audio lag so that any decline in the facial overshadowing effect could not be attributable to mere habituation to a pre-exposed face, as shown by Cook and Wilding (2001). [↑](#footnote-ref-1)
2. One may consider that ceiling effects in accuracy and d’ may have compromised the ability to reveal greater overshadowing under asynchronous than synchronous conditions. However, when reaction time data were considered, there was no pattern of facial overshadowing at all. These data are not reported further given variance in the accuracy and d’ data. [↑](#footnote-ref-2)