The effect of distraction on face and voice recognition.

Sarah V Stevenage,* Greg J Neil, Jess Barlow, Amy Dyson,

Catherine Eaton-Brown & Beth Parsons

Psychology, University of Southampton

CORRESPONDENCE may be sent to: Dr Sarah Stevenage, Psychology, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. Tel: 02380 592234; Fax: 02380 594597; email: svs1@soton.ac.uk

The final publication is available at Springer via http://dx.doi.org/ 10.1007/s00426-012-0450-z

Abstract

The results of two experiments are presented which explore the effect of distractor items on face and voice recognition. Following from the suggestion that voice processing is relatively weak compared to face processing, it was anticipated that voice recognition would be more affected by the presentation of distractor items between study and test compared to face recognition. Using a sequential matching task with a fixed interval between study and test that either incorporated distractor items or did not, the results supported our prediction. Face recognition remained strong irrespective of the number of distractor items between study and test. In contrast, voice recognition was significantly impaired by the presence of distractor items regardless of their number (Experiment 1). This pattern remained whether distractor items were highly similar to the targets or not (Experiment 2). These results offer support for the proposal that voice processing is a relatively vulnerable method of identification.

The effect of distraction on face and voice recognition.

Several studies are emerging with a focus on the voice as a means of person recognition. Across a number of these studies, results suggest that whilst face recognition proceeds with speed, accuracy and confidence, voice recognition is achieved more slowly, shows more errors, and is completed with less confidence (see Yarmey, 1995 for a review). At a theoretical level, these findings have been used to suggest that voice recognition is represented by a weaker pathway than face recognition. As a consequence, it may be anticipated that voice recognition will be more vulnerable to interference than face recognition. The present paper reports on two experiments designed to test this prediction.

The relative weakness of voices

The suggestion that voices represent a relatively weak route to person recognition rests on a growing literature. A seminal paper is provided by Hanley, Smith and Hadfield (1998) who showed a greater incidence of ‘familiar only’ states when listening to a familiar voice than when seeing a familiar face (see also Ellis, Jones & Mosdell, 1997; Hanley & Turner, 2000). In fact, the recognition of a familiar person from voice and face could only be equated if the face was presented as a substantially blurred image (Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009). Furthermore, this result held whether publicly familiar or personally familiar targets were used as stimuli (see Barsics & Brédart, 2011).

Added to this, the retrieval of both episodic and semantic information has been shown to be more difficult when cued by a voice than by a face. For instance, the retrieval of a particular recollective instance (‘Remember’ state) is relatively rare, whereas a sense of familiarity without recollection of an instance (‘Know’ state) is more likely, when cued with a voice rather than a face (Barsics & Brédart, 2011), even when the face is again blurred (Damjanovic & Hanley, 2007). Similarly, the retrieval of a particular piece of semantic information such as an occupation (Hanley & Damjanovic, 2009; Hanley, Smith & Hadfield, 1998), or a topic taught by a school teacher (Barsics & Brédart, 2011; Brédart, Barsics & Hanley, 2009), is again easier when cued by the (blurred) face than by the voice. All of these findings suggest that the voices are relatively weak, both in triggering recognition, and in enabling subsequent access to semantic, episodic, or name information.

Finally, it is worth noting the results of two priming studies. Using a repetition priming paradigm, both Schweinberger, Herholz and Stief (1997) and Stevenage, Hugill and Lewis (2012) suggested that whilst faces can prime subsequent voice recognition, voices do not prime subsequent face recognition. This asymmetry is important. It again signals the voice as a weaker input to person recognition than the face, hence its reduced capacity as a prime. These data converge with the above findings. Consequently, with voices identified as a weak recognition route, the suggestion that follows is that they may be more vulnerable than faces to interference.

Interference Effects

When considering interference effects, several studies are of relevance. First, it is useful to consider the effects of delay on face and voice recognition. In this regard, it is notable that face recognition appears remarkably robust over time. One of the best cited examples is provided by Bahrick, Bahrick and Wittlenger (1975) who observed 90% recognition of former class mates from yearbook photographs, even across a delay of up to 50 years. In contrast, voice recognition appears to be substantially compromised over a relatively short delay. For example, whilst voice recognition remained unimpaired across a delay of 24 hours (Clifford, 1980; Saslove & Yarmey, 1980), it showed significant decline after just three weeks between study and test (Kersholt, Jansen, Van Amelsvoort & Broeders, 2006), and hit rates deteriorated from 83% at immediate testing, to only 35% after a three month delay (McGehee, 1937). Whilst these effects are striking, their interpretation is unclear because it is not possible within these studies to distinguish the effects of delay from the effects of interference during the intervening period.

A second series of studies are valuable in this regard. These hold delay relatively constant by controlling the length of time between study and test. The result is the ability to examine interference effects in isolation. In this regard, it is worth noting that face recognition is not immune to interference effects, and an elegant demonstration of this is provided by Hinz and Pezdek (2001). In this paper, participants were asked to study a target face for 60 seconds. One week later, they viewed an intervening lineup, and two days afterwards, they engaged in a recognition test. The recognition test consisted of a 6 person lineup in which the target was either present or absent. In addition, however, a critical lure from the intervening lineup was also either present or absent. Performance was significantly affected by this intervening lure in that the capacity to recognise the target fell significantly when the lure was also present, and the false recognition of the lure rose significantly when the target was absent. These data suggest that face recognition can be substantially affected by the presentation of other (lure) faces between study and test.

Interference effects have also been shown for voice recognition, and so far, studies have reported on cross-modality interference rather than the within-modality interference explored above. In this regard, findings suggest that the co-presentation of a face alongside a voice at study (McAllister, Dale, Bregman, McCabe & Cotton, 1993; Stevenage, Howland & Tippelt, 2011) and at test (Cook & Wilding, 1997), leads to an impairment in subsequent voice recognition. Importantly, however, and in common with the priming results above, these interference effects are not symmetrical: whilst the co-presentation of the face affected subsequent voice recognition, the co-presentation of the voice did not affect subsequent face recognition (Stevenage et al., 2011). This asymmetry has been described as the *face* *overshadowing effect*, suggesting again a relative vulnerability of the voice compared to the face.

What remains to be seen, however, is whether voice recognition is more susceptible than face recognition to within-modality interference as well as cross-modality interference. An answer to this question enables scrutiny of the ‘weaker voice’ thesis in a way that holds delay constant, and respects any imbalance across modalities as revealed in the face overshadowing effect. Consequently the question being asked is whether voice recognition is more affected by distractor voices than face recognition is affected by distractor faces. This provides a strong test of the hypothesis and if voice recognition is indeed represented by a weaker pathway than face recognition, it may be anticipated that it will be impaired to a greater degree by the presence of intervening distractors even in this controlled test.

Method

Design

A 2 x 3 mixed design was used in which the type of stimuli (face, voice) was varied between participants, and the number of distractor items (0, 2, 4) was varied within participants. A sequential matching task was used, with study and test phases separated by a fixed length intervening period during which distractor items were presented. Accuracy of same/different response, together with self-rated confidence, represented the dependent variables.

Participants

A total of 40 participants (15 males, 25 females) took part in the present study either on a volunteer basis or in return for course credit. All had normal or corrected-to-normal hearing and vision, and all were unfamiliar with the faces or voices used. Participants were randomly assigned to complete either the face recognition task (n = 20; 12 females) or the voice recognition task (n = 20; 13 females), and groups were matched for age (face: mean age = 24.6 (SD = 5.71); voice: mean age = 27.0 (SD = 6.47); $t_{(38)} = 1.24$, *ns*).

Materials

Faces: A total of 42 male and 42 female faces were used in the face recognition test. These were obtained from a student record database, and all individuals had consented to the use of their face within this research. All stimuli were selected to be free from distinctive features such as moles, facial hair, scars or spectacles and all were depicted with a natural neutral or smiling expression. Across trials, 24 males and 24 females were designated as targets, and the identity of these was counterbalanced across participants. A further 6 male and 6 female faces were designated as distractor items and were presented in the intervening phase only. Finally, 12 male and 12 female faces were designated as lures, and were presented at the test phase of 'different' trials. All face stimuli were prepared within Corel Photopaint to be extracted from their background, standardised for size based on inter-pupillary distance, and presented as greyscale images mounted within a white square measuring 7 x 7 cm. Within this frame, the face measured approximately 3 cm wide x 4 cm high.

Voices: As with the face trials, a total of 42 male and 42 female voice clips were used in the voice recognition test. These were obtained by asking student-aged speakers to utter a standard 15 word phrase designed to maintain attention without being threatening ('I think the most important thing to remember is to keep calm and stay safe'). All voice stimuli were prepared within Audacity 3.1 to extract the speech segment from any pauses before, during, or after the voiced statement resulting in clips that had a mean length of 4.58 seconds (SD = .73; min = 3.00, max = 6.00).

Stimuli were presented, and data were recorded using Superlab 2.1, running on a Latitude E6400 laptop PC with a 14" colour monitor and a screen resolution of 1400 x 990 pixels. Voices were presented via headphones that covered the ear ensuring audibility of stimuli and minimising acoustic interference.

Procedure

Participants were tested individually within a quiet testing environment. A sequential matching task was used in which participants experienced a study phase, an intervening period, and a test phase. At test, they were asked to determine whether a stimulus was the 'same' or 'different' to the target presented at study. Responses were made by pressing 'S' for 'same', and 'D' for 'different' as quickly but as accurately as possible.

Practice was provided through 20 trials with the words 'same' and 'different', enabling participants to map their responses to the correct keys. Following this, 12 further practice trials used written nouns as target and test stimuli, enabling participants to become familiar with the trial format. In these trials, participants were presented with a written target word (i.e., child) followed by either 0, 2, or 4 distractor words, before being presented with either the target (child) again, or a lure (i.e., train). Participants experienced four trials at each level of distraction, blocked according to the number of distractor items. Data were recorded but were not analysed.

The main trials took a similar format whether face recognition or voice recognition was being tested. The study phase was introduced with a 'next trial' prompt for 250 ms. Participants were then presented with either a face or a voice for a fixed exposure duration of 750 ms (faces) or 4 seconds (voices), and these exposure durations were selected to encourage adequate encoding of both the face and the voice, with the latter unfolding over time. Participants were asked to give a rating of distinctiveness for the face or voice using a scale of 1-7 (where 1 = 'very typical' and 7 = 'very distinct'). This encouraged attention to the target.

A fixed intervening period of 16 seconds followed, during which participants were either presented with 0, 2, or 4 distractor items, each of which remained visible or audible for a 4 second period. Thus, 4 distractors in sequence lasted the full intervening period of 16 seconds, whereas 2 distractors in sequence lasted for 8 seconds of the intervening period. Trials were presented in a random order, but were organised in blocks according to the number of distractors presented. In order to avoid simple fatigue effects, the order of blocks was counterbalanced across participants to provide either increasing (0, 2, 4) or decreasing (4, 2, 0) difficulty. Care was also taken to ensure that distractors were of the same stimulus type as the targets, hence faces were distracted with faces, and voices were distracted with voices. Moreover the gender of the targets and distractors was matched, ensuring that the similarity between target and distractors was optimised. The task at this stage was to attend to these stimuli (or to the blank screen) but not to respond.

In the test phase, participants were presented with a single stimulus, and their task was to determine whether the target and test were the 'same' or 'different'. Accuracy was recorded, together with self-rated confidence using a scale of 1-7 (where 1 = 'not at all confident' and 7 = 'highly confident'). The entire process lasted approximately 30 minutes after which participants were thanked and debriefed.

Results

Sensitivity of Discrimination

Accuracy on the matching task was manipulated to give a measure of sensitivity of discrimination (d'). In contrast to an analysis of accuracy, this measure is free from the effect of response bias[1]. Table 1 summarises performance for face recognition and voice recognition across the three levels of distraction, and a 2 x 3 mixed Analysis of Variance

---

[1] Analysis in signal detection terms enables scrutiny of both sensitivity of discrimination (d') and bias (C). Analysis of bias here revealed no effects either of the number of intervening stimuli ($F_{(2, 76)} < 1$, *ns*), stimulus type ($F_{(1, 38)} < 1$, *ns*) or their interaction ($F_{(2, 76)} < 1$, *ns*).

(ANOVA) was used to examine the impact of each variable in isolation and in combination. This revealed a significant main effect of stimulus type ($F_{(1, 38)} = 20.56$, $p < .001$, $ή^2 = .36$) with performance being better for faces than voices. In addition, there was a main effect of the number of distractors ($F_{(2, 76)} = 21.20$, $p <.001$, $ή^2 = .36$) with performance becoming increasingly worse as the number of distractors increased. Both effects were qualified by a large and expected interaction between stimulus type and distraction ($F_{(2, 75)} = 10.11$, $p < .001$, $ή^2 = .21$). Post-hoc contrasts confirmed this to be due to robust performance as distraction increased when recognising faces ($F_{(2, 38)} = 1.34$, *ns*), but a clear and significant decline in performance as distraction increased when recognising voices ($F_{(2, 38)} = 26.25$, $p < .001$, $ή^2 = .58$). This latter effect was explained by a significant decline in performance when presented with zero distractors compared to *any* distractors ($F_{(1, 19)} = 46.48$, $p < .001$, $ή^2 = .71$) but no further decline in performance when increasing from two to four distractors ($F_{(1, 19)} = 3.18$, *ns*).

(Please insert Table 1 about here)

Confidence in Correct Decisions

Due to the small number of incorrect decisions in some conditions, self-rated confidence was examined for correct decisions only, and is summarised in Table 2 for face and voice recognition across the different levels of distraction, and across ‘same’ and ‘different’ trials separately. A 3 way mixed ANOVA explored the effects of stimulus type, distraction, and trial-type. This revealed a main effect of stimulus type ($F_{(1, 38)} = 15.72$, $p < .001$, $ή^2 = .29$) with confidence being higher when recognising faces than when recognising voices. In addition, there was a main effect of distraction ($F_{(2, 76)} = 29.52$, $p < .001$, $ή^2 = .44$) with confidence being highest when there were zero distractors, and lowest when there were four distractors. Finally, there was a main effect of trial type ($F_{(1, 38)} = 20.93$, $p < .001$, $ή^2 = .35$) with confidence being higher in ‘different’ trials than in ‘same’ trials. Interactions emerged between distraction and stimulus type ($F_{(2, 76)}$ = 5.38, p < .01, ή^2 = .12), and between all three factors ($F_{(2, 76)}$ = 3.27, p < .05, ή^2 = .08).

(Please insert Table 2 about here)

Analysis of the simple main effects revealed that there was significant interference for same trials and for different trials when tested with both faces and voices (all Fs $_{(2, 38)}$ > 5.15, p < .02, ή^2 = .21). All effects revealed a drop in confidence between zero distractors and *any* distractors (all $Fs_{(1, 19)}$ > 7.46, p < .025, ή^2 > .282), but no further decrease in confidence between two and four distractors (all Fs$_{(1, 19)}$ < 4.32, *ns*). Additional analysis, however, revealed that when alpha was adjusted to account for two post-hoc comparisons, the *magnitude* of this interference effect (confidence when zero distractors – confidence when four distractors) was greater for voice recognition than for face recognition on same trials (face = .45 (SD = .74); voice = 1.17, (SD = 1.05); $t_{(38)}$ = 2.52, p < .016) but not on different trials (face = .33 (SD = .40); voice = .74 (SD = .74); $t_{(38)}$ = 2.20, p > .025). This was most likely due to the large and negative effect of distraction on confidence in voice recognition during same trials in particular.

Discussion

The results of Experiment 1 confirmed the expectation that face recognition would remain robust despite short term distraction. In contrast, and again in line with expectation, voice recognition showed a significant and substantial decline in performance as the level of distraction increased. This impairment was demonstrated both through a behavioural measure of performance (d') and through a metamemory measure (confidence). More particularly, both sensitivity of discrimination, and confidence showed a significant decline as soon as *any* distraction was introduced, but the extent of distraction (2 or 4 items) did not matter. These results supported the view that voice recognition would be weaker and hence more vulnerable to distraction compared to face recognition. Several aspects of the current results are, however, worthy of further consideration.

First, it was notable that voice recognition was equivalent to face recognition in the baseline condition when no distractors were presented ($t_{(38)}$ = < 1, *ns*). At some level, this was surprising because it might have been expected that voice recognition would be worse than face recognition even in this most optimal condition. Remember, however, that the sequential matching task was quite unlike either an old/new recognition task or a lineup task used in previous research. Both of these tasks involve the presentation of a set of study and/or test stimuli prior to performance and both thus involve a time delay and a distraction. In contrast, the sequential matching task with zero distractors provided a constant delay between study and test, but provided no distraction at all. Equivalent performance on the face and the voice recognition task here becomes important because it suggests that mere effects of delay (here over 16 seconds) are equivalent for face and voice recognition. In this regard, previous deficits in voice recognition compared to face recognition may be attributable to the differential impact of distraction for voice and face recognition.

Second, it was notable that participants were able to reflect accurately on their performance in the face and voice recognition tasks. In this regard, when performance was good, confidence was high, and when performance declined in the voice recognition task, confidence declined also. This successful calibration emerged to a significant degree for both faces and voices and across both ‘same’ and ‘different’ trials. However, the effect was largest when recognising voices in ‘same’ trials. This fits with the observation that accuracy fell most substantially for voice recognition in ‘same’ trials as the number of distractor stimuli rose. Participants appeared to have good awareness of this.

In accounting for these results, it may be reasonable to consider whether voice recognition will always be so substantially affected by distraction, or whether the magnitude of effect revealed here is dependent on the type of distraction presented. If the underlying premise is that the voice recognition impairment here is indicative of a weak voice recognition pathway, then this impairment may be anticipated no matter what type of distraction is provided. Experiment 2 provides a test of this hypothesis through explicit manipulation of the strength of the distractor.

Experiment 2

The results of the previous experiment demonstrated a clear vulnerability to distraction when recognising voices, but clear resilience when recognising faces. However, it was unclear whether voice recognition would be impaired regardless of the type of distraction, or was impaired here because the distractors were so similar in form to the target. In order to explore this issue further, Experiment 1 was replicated here using distractors that were either similar to, or different from, the target stimuli. Specifically, strong interference was represented by using *gender-matched* distractors for a given target (i.e., 0, 2, or 4 female voices were distractors for the female voice targets). This condition replicates Experiment 1. In contrast, weak interference was represented by using gender-opposite distractors for a given target (i.e., 0, 2, or 4 female voices were distractors for the male voice targets). The rationale for this manipulation was that if the previous effect emerged as a result of the strength of the distractors, then weakening the distractors may weaken the effect. If, however, the previous effect emerged as a result of the weakness of the target per se, then weakening the distractors would have no effect.

Method

Design

A 2 x 3 x 2 mixed design was used in which stimulus type (face, voice) was varied between participants, and the number of distractors (0, 2, 4) was varied within participants as above. In addition, distractor strength was systematically varied between participants to provide strong distractors and weak distractors. As before, a sequential matching task was used, with study and test phases separated by a fixed length intervening period. Accuracy of same/different response, together with self-rated confidence, represented the dependent variables.

Participants

A total of 65 participants (8 males, 57 females, mean age = 20.48 years (SD = 5.02)) took part in the present study either on a volunteer basis or in return for course credit. All had normal or corrected-to-normal hearing and vision, and were unfamiliar with the stimuli used. In addition, none had taken part in Experiment 1. Participants were randomly assigned to the face recognition task (n = 33, 27 females) or the voice recognition task (n = 32, 30 females), and groups were matched for age and gender (face: mean age = 19.55 (SD = 2.46); voice: mean age = 21.22 (SD = 6.68); $t_{(63)} = 1.35$, *ns*).

Materials

The materials were identical to those used in Experiment 1.

Procedure

The procedure was identical to Experiment 1 with the exception that those participants in the strong distractor condition were presented with female faces or voices as distractors for female targets, and male faces or voices as distractors for male targets. In contrast, and as an extension to Experiment 1, those participants in the weak distractor condition were presented with female faces or voices as distractors for male targets, and male faces or voices as distractors for female targets. All other aspects of the procedure were as described above.

Results

Sensitivity of Discrimination

As in Experiment 1, accuracy of performance was manipulated to give a measure of sensitivity of discrimination (d') and this is summarised in Table 3. A 2 x 3 x 2 mixed ANOVA was used to explore the effects of stimulus type, distraction and distractor strength on recognition performance[2]. This revealed a significant main effect of stimulus type ($F_{(1, 61)} = 22.18$, $p < .001$, $\acute{\eta}^2 = .27$) with performance being better for face recognition than voice recognition. In addition, there was a significant main effect of distraction ($F_{(2, 122)} = 17.54$, $p < .001$, $\acute{\eta}^2 = .22$) with performance being best when there were zero distractors, and worst when there were four distractors. Finally, a significant interaction between stimulus type and distraction emerged ($F_{(2, 122)} = 7.33$, $p < .001$, $\acute{\eta}^2 = .11$). Notably, there was no effect of distractor strength either alone ($F_{(1, 61)} = 2.02$, $p > .05$), or in combination with any other variable(s) (all *F*s < 1.24, $p > .05$).

(Please insert Table 3 about here)

Exploration of the significant interaction between stimulus type and distraction reiterated the results of Experiment 1. Specifically, post-hoc analysis revealed no influence of distraction on face recognition ($F_{(2, 64)} = 2.46$, $p > .05$) confirming robust recognition regardless of the level of distraction provided. In contrast, there was a significant main effect of distraction on voice recognition ($F_{(2, 62)} = 21.21$, $p < .001$, $\acute{\eta}^2 = .41$), with a clear decline in performance between zero distraction and *any* distraction ($F_{(1, 31)} = 30.67$, $p < .001$, $\acute{\eta}^2 = .35$), and a small but significant further decline as distraction was increased from 2 to 4 items ($F_{(1, 31)} = 6.58$, $p < .025$, $\acute{\eta}^2 = .18$). These findings confirmed the effect of distraction on voice

---

[2] As in Experiment 1, analysis of bias is included here to explore the effect of stimulus type, distraction and distractor strength on responding. Only one interaction emerged as a weak effect (distraction x stimulus type: $F_{(2, 122)} = 3.23$, $p < .05$, $\acute{\eta}^2 = .05$). However, there was no effect of distraction on bias for faces ($F_{(2, 64)} = 1.17$, *ns*) or voices ($F_{(2, 62)} = 2.53$, *ns*), and there was no significant difference in levels of bias between faces and voices at either zero distractors ($t_{(63)} = -1.53$, *ns*), two distractors ($t_{(63)} = 1.20$, *ns*) or four distractors ($t_{(63)} = -1.61$, *ns*). This interaction instead seemed to capture a reversal of small levels of bias between faces and voices when there were two distractors. No other main effects of interactions reached significance (*F*s < 1.38, *ns*).

recognition but not face recognition. Most importantly, distractor strength did not affect its capacity to impair performance in any way.

Confidence in Correct Decisions

As in Experiment 1, self-rated confidence was examined for correct decisions only. Table 4 summarises these data, and a 2 x 2 x 3 x 2 mixed ANOVA was used to examine the effects of stimulus type, distractor strength, level of distraction, and trial type respectively. This revealed significant main effects for all variables except distractor strength. More specifically, there was a main effect of stimulus type ($F_{(1, 61)} = 25.11$, $p < .001$, $ή^2 = .29$) with greater confidence in face recognition than voice recognition. Similarly, there was a significant effect of distraction ($F_{(2, 122)} = 25.15$, $p < .001$, $ή^2 = .29$), with greater confidence when there were zero distractors, and least confidence when there were four distractors. There was also a significant effect of trial type ($F_{(1, 61)} = 10.31$, $p < .005$, $ή^2 = .15$), with greater confidence on 'different' trials than on 'same' trials. These effects replicated those found in Experiment 1.

(Please insert Table 4 about here)

In addition, there was a significant interaction between level of distraction and stimulus type ($F_{(2, 122)} = 5.33$, $p < .01$, $ή^2 = .08$). Post hoc examination was conducted through two repeated-measures ANOVAs to examine confidence separately for face and voice recognition. These revealed a significant distraction effect on self rated confidence both for face recognition ($F_{(2, 62)} = 10.34$, $p < .001$, $ή^2 = .24$) and voice recognition ($F_{(2, 64)} = 16.22$, $p < .001$, $ή^2 = 34$). However, consideration of the *magnitude* of this distraction effect showed there to be a greater decline in confidence for voice recognition (decline = .70 (SD = .81)) than for face recognition (decline = .31, (SD = .47), ($t_{(63)} = 2.39$, $p < .025$).

Discussion

The results of Experiment 2 reiterated in all regards the findings of Experiment 1. Specifically, face recognition remained robust across increasing levels of distraction, but voice recognition declined significantly as distraction rose. This was shown both through the behavioural measure (d') and the metamemory measure (confidence) as before.

Some small differences in results exist between Experiments 1 and 2, and these may reflect a somewhat more difficult set of conditions overall in Experiment 2. For example, as well as a reduction in d' from zero distractor items to *any* distractor items when recognising voices, there was a small but significant decrease in performance as distraction rose further from 2 to 4 distractor items. In addition, the analysis of confidence revealed a drop in confidence across distraction when recognising voices in both 'same' and 'different' trials, rather than just in the 'same' trials of Experiment 1. Both of these variations in results may indicate a slight increase in difficulty felt by participants as a whole in Experiment 2. However, it should be reiterated that there was no influence of distractor strength (strong, weak) either alone or in combination with any other variable. As such, the effect of interference can confidently be attributed to the weakness of the voice target rather than to the strength of the voice distractor.

General Discussion

The results presented across two studies have demonstrated and confirmed robust face recognition but impaired voice recognition as distraction increases. Moreover, confidence declined more with increasing distraction when recognising voices than when recognising faces, and this may impact on the likelihood of volunteering a recognition decision in a more realistic scenario. The design of both studies precludes explanation on the basis of the mere passage of time, as the intervening period was held constant and yet differential effects emerged. Moreover, the results cannot be attributed to the strength of distraction, as the effects were demonstrable whether strong or weak distractors were used. In this sense, the relative vulnerability of voice recognition to within-modality interference was likely to be the result of a relatively weak voice recognition pathway per se.

It was perhaps surprising that face recognition in both experiments showed no decline as the level of distraction was increased. Based on the results of Hinz and Pezdek (2001), an effect of interference might have been anticipated. In this regard, the current demonstrations gain rigour by being demonstrated across multiple targets whereas the results of Hinz and Pezdek are based on the recognition of a single target. This leaves the latter result open to potential item effects. In addition, the use of a lineup during both the intervening period and the test may have induced transfer-appropriate-processing in Hinz and Pezdek's study, magnifying the influence of the intervening stimuli on their final test. Indeed, the critical lure was in the same lineup position (position 5) at both the distraction and the test stage, maximising the potential for episodic effects to reduce recognition performance. These issues are avoided in the current paper, although the ecological validity of Hinz and Pezdek's study is a point that will be returned to later.

In accounting for these results, we return to the notion of the voice pathway as being relatively weak compared to the face pathway in a person recognition system. This is articulated in several publications (Barsics & Brédart, 2011, 2012; Damjanovic & Hanley, 2007; Ellis et al., 1997; Hanley & Damjanovic, 2009; Hanley, et al., 1998; Hanley & Turner, 2000; Schweinberger et al., 1997; Stevenage et al., 2012) and is supported by a considerable empirical literature. In this regard, it is our contention that the greater susceptibility to distraction for voices exists because of a relative inability to form a strong representation to underpin a robust subsequent recognition. This by itself may not be sufficient as an explanation though, and here we provide discussion which may *account* for the relative weakness of voice processing rather than merely assuming and relying on it.

Three mechanisms are considered here to account for the relative weakness of voices compared to faces, but all recognise that relative voice weakness stems from a failure to differentiate one person from another to the same extent from their voice as from their face. This may emerge because we have (i) less experience with voices than with faces, (ii) a reduced need to process identity from voices than from faces, or (iii) less expertise in structurally encoding voices than faces. In evaluating these mechanisms, there is perhaps only weak evidence in favour of the first suggestion. In particular, there may well be less experience of voices than faces when considering celebrity targets, but when personally familiar targets are considered, this differential is presumably reduced. Importantly however, voice recognition remains weaker than face recognition whether personally or publicly familiar stimuli are used (see Barsics & Brédart, 2011). Moreover, in a more recent demonstration, Barsics and Brédart (2012) used a learning paradigm with pre-experimentally unfamiliar faces and voices so that the exposure to each stimulus type could be carefully controlled. Even under these conditions, voice recognition remained significantly weaker than face recognition for the recall of occupations (Experiment 1) and occupations and names (Experiment 2). These data cast doubt on an explanation based on differential experience between faces and voices at a stimulus specific level. Nevertheless, as our experience with all faces and voices shapes our overarching skills, the point is included here.

In contrast, there is good intuitive support for the second and third mechanisms. With voices often accompanying faces, the voice does not need to signal identity and instead may often be processed in terms of the message rather than the messenger (Stevenage, Hugill & Lewis, 2012). Similarly, and perhaps consequently, we may lack the ability to encode a voice, or to differentiate one voice from its competitor, with the same level of expertise as for faces. The result is the relatively weak activation of a voice recognition unit compared to a face recognition unit, with the result that associations or links from the voice recognition unit to the PIN and beyond are similarly weak. Hanley and colleagues provide robust empirical demonstration of the consequence of these weak links through a greater difficulty in retrieval of semantic, episodic, and name information when cued by the voice than by the face.

Regardless of its cause, the *consequence* of a relatively weak pathway for voice recognition can then be articulated. When a voice is heard, the associated PIN may only be weakly activated. In structural terms, it then has less capacity to withstand any inhibitory effects from distractor items, and it has less capacity to be re-activated at test through a self-priming mechanism (see Burton, Bruce & Johnston, 1990). It also has less capacity to receive back-activation from any semantic, episodic, or name information because of the reduced likelihood of these being associated or retrieved. The result demonstrated here is that a weaker voice recognition route is more affected by distractors than a stronger face recognition route.

The implications of these results for the police or court setting are clear. Voice recognition is vulnerable to interference and thus may not meet the level of evidentiary standard required of courtroom evidence. Before drawing this strong conclusion, there would, however, be merit in exploring interference effects in face and voice recognition using a more ecologically valid method such as that presented by Hinz and Pezdek (2001). If convergent results emerge, and voice recognition remains more vulnerable to interference under this more realistic method, then serious questions would need to be asked regarding the future of voice recognition within police investigations and court proceedings.

References

Bahrick, H.P., Bahrick, O.O. & Wittlenger, R.P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54-75.

Barsics, C., & Brédart, S. (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition, 20(2),* 303-308.

Barsics, C., & Brédart, S. (2012). Recalling semantic information about newly learned faces and voices. *Memory, 20(5),* 527-534.

Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology, 21*, 1013-1021.

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

Clifford, B.R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior, 4(4),* 373-394.

Cook, S., & Wilding, J. (1997). Earwitness Testimony 2: Voices, Faces and Context. *Applied Cognitive Psychology, 11,* 527-541.

Damjanovic, L. (2011). The face advantage in recalling episodic information: Implications for modelling human memory. *Consciousness and Cognition, 20(2),* 309-311.

Damjanovic, L., & Hanley, J.R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory and Cognition, 35*, 1205-1210.

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*, 143-156.

Hanley, J.R., & Damjanovic, L. (2009). It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory, 17*, 830-839.

Hanley, J.R., Smith, S.T., & Hadfield, J. (1998). I recognise you but can't place you. An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology, 51A(1),* 179-195.

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.

Hinz, T., & Pezdek K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior, 25,* 185-198.

Kerstholt, J.H, Jansen, N.H.M., van Amelsvoort, A.G., & Broeders, A.P.A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology, 20,* 187-197.

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.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of GeneralPsychology, 17,* 249-271.

Saslove, H., & Yarmey, A.D. (1980). Long-term auditory memory; Speaker identification. *Journal of Applied Psychology, 65,* 111-116.

Schweinberger, S.R., Herholz, A., & Stief, V. (1997). Auditory Long-term Memory: Repetition Priming of Voice Recognition. *Quarterly Journal of Experimental Psychology, 50A(3),* 498-517.

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

Stevenage, S.V., Hugill, A.R., & Lewis, H.G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, in press

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

Table 1: Sensitivity of discrimination, together with accuracy levels, (and standard deviation) for face and voice recognition under increasing levels of distraction in Experiment 1.

	0 distractors	2 distractors	4 distractors
FACE RECOGNITION			
Sensitivity of discrimination (d')	4.22 (.82)	4.03 (.73)	3.82 (.96)
Proportion accuracy for SAME trials	.96 (.07)	.94 (.10)	.93 (.09)
Proportion accuracy for DIFFERENT trials	.99 (.04)	.98 (.06)	.97 (.07)
VOICE RECOGNITION			
Sensitivity of discrimination (d')	4.16 (.73)	2.67 (1.22)	2.18 (1.28)
Proportion accuracy for SAME trials	.94 (.12)	.79 (.21)	.78 (.17)
Proportion accuracy for DIFFERENT trials	.99 (.04)	.89 (.16)	.84 (.12)

Table 2: Self-rated confidence in correct decisions for face and voice recognition under increasing levels of distraction in Experiment 1. (Ratings are made out of 7, where 1 = 'not at all confident' and 7 = 'very confident indeed'. Standard Deviation is provided in parentheses.)

		0 distractors	2 distractors	4 distractors
FACE RECOGNITION				
	SAME trials	6.36 (.79)	6.29 (.88)	5.91 (.97)
	DIFFERENT trials	6.78 (.30)	6.29 (.86)	6.45 (.59)
VOICE RECOGNITION				
	SAME trials	5.98 (.90)	5.08 (1.00)	4.81 (1.00)
	DIFFERENT trials	6.13 (.65)	5.59 (.96)	5.40 (.90)

Table 3: Sensitivity of discrimination, together with accuracy levels, (and standard deviation) for face and voice recognition under increasing levels of strong and weak distraction in Experiment 2.

	0 distractors	2 distractors	4 distractors
FACE RECOGNITION			
Strong (gender-matched) distraction:			
Sensitivity of discrimination (d’)	4.09 (.72)	3.76 (.97)	3.77 (1.08)
Prop accuracy for SAME trials	.93 (.11)	.94 (.11)	.91 (.13)
Prop accuracy for DIFFERENT trials	1.00 (.00)	.94 (.09)	.97 (.05)
Weak (gender-mismatched) distraction:			
Sensitivity of discrimination (d’)	4.17 (.75)	3.67 (1.00)	4. 01 (.91)
Prop accuracy for SAME trials	.93 (.11)	.88 (.17)	.91 (.14)
Prop accuracy for DIFFERENT trial	1.00 (.00)	.98 (.05)	.99 (.03)
VOICE RECOGNITION			
Strong (gender-matched) distraction:			
Sensitivity of discrimination (d’)	3.60 (.85)	2.48 (1.07)	2.24 (1.29)
Prop accuracy for SAME trials	.91 (.13)	.77 (.20)	.82 (.19)
Prop accuracy for DIFFERENT trials	.95 (.08)	.91 (.10)	.80 (.15)
Weak (gender-mismatched) distraction:			
Sensitivity of discrimination (d’)	3.83 (1.12)	3.28 (.93)	2.61 (1.20)
Prop accuracy for SAME trials	.93 (.12)	.87 (.13)	.83 (.16)
Prop accuracy for DIFFERENT trials	.95 (.10)	.95 (.06)	.86 (19)

Table 4: Self-rated confidence in correct decisions for face and voice recognition under increasing levels of strong and weak distraction in Experiment 2. (Ratings are made out of 7, where 1 = ‘not at all confident’ and 7 = ‘very confident indeed’. Standard deviation is provided in parentheses.)

	0 distractors	2 distractors	4 distractors
FACE RECOGNITION			
Strong (gender-matched) distraction:			
SAME trials	6.52 (.65)	6.34 (.63)	6.04 (.73)
DIFFERENT trials	6.77 (.37)	6.51 (.62)	6.23 (.75)
Weak (gender-mismatched) distraction:			
SAME trials	6.17 (1.13)	6.17 (1.15)	6.17 (1.20)
DIFFERENT trials	6.43 (1.16)	6.25 (1.24)	6.25 (1.10)
VOICE RECOGNITION			
Strong (gender-matched) distraction:			
SAME trials	5.53 (.81)	5.26 (1.08)	4.97 (1.12)
DIFFERENT trials	6.09 (.85)	5.27 (1.12)	5.37 (1.07)
Weak (gender-mismatched) distraction:			
SAME trials	5.59 (.79)	4.95 (.94)	5.01 (.94)
DIFFERENT trials	5.97 (.67)	5.35 (.87)	5.03 (.78)