Psychologica Belgica

Stevenage, S. V. and 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

INVITED REVIEW

Hearing Faces and Seeing Voices: The Integration and Interaction of Face and Voice Processing

Sarah V Stevenage* and Greg J Neil*

Cognitive understanding of voice recognition has borrowed much from the area of face processing, both in terms of the theoretical framework within which results are interpreted, and the methodology used to assess performance. A considerable body of research now exists to suggest that voice recognition may proceed in parallel with face recognition, and that the two pathways may combine to inform person recognition. However, rather than being independent or equivalent, these parallel pathways appear to interact to reveal interesting interference effects. The present paper reviews a series of studies that focus on a considerable and growing literature. The vulnerability of voice processing will be explored relative to face processing, and the interaction of these two pathways will be examined with reference to broader theoretical frameworks for person recognition.

Keywords: voice processing; face processing; identification; integration of processes

The capacity to identify someone from their face is relatively well-researched and understood at both a behavioural and neuropsychological level. Levels of performance are impressively strong, and neuropsychological mechanisms highlight particular areas of the right temporal lobe which underpin this performance. Against this backdrop, identification from a voice is now receiving a high level of attention, perhaps with the hope of revealing similarly impressive levels of performance. Indeed, evidence of parallel neural mechanisms in humans and macaques (see Yovel & Belin, 2013 for a review), and evidence of voice recognition capacities in newborn infants (see Beauchemin et al., 2011) suggest that voice processing is a phylogenetically ancient process to which we bring some innate predisposition or neurological preparedness. These demonstrations combine with a surge of experimental and neuropsychological interest in voice processing to explore the parallels that may exist between face recognition and voice recognition (see Brédart & Barsics, 2012; Hanley, 2014; Latinus & Belin, 2011; Yovel & Belin, 2013 for useful overviews).

Surprisingly, however, research has consistently reported voice recognition to be relatively weak compared to face recognition. The current paper reflects on our capacity

* Department of Psychology, University of Southampton, UK
svs1@soton.ac.uk

Corresponding Author: Sarah V Stevenage

to recognise voices, and explores the interaction and integration of voice recognition with face recognition when recognising an individual. Our view, however, is that merely focussing on the voice as a cue to identity is akin to looking at only one piece of a larger puzzle. Consequently, we take the opportunity to expand our discussion, setting voice recognition in the context of a wider set of tasks to which the voice can contribute.

A Theoretical and Neural Framework for Voice Recognition

The starting point for many researchers when considering voice recognition has been the seminal paper by Bruce and Young (1986) on face recognition. They presented an Information Processing Model in which familiar face recognition was achieved through initial structural processing of the face resulting in a view- and expression-invariant representation of an individual. If the individual was known, a Face Recognition Unit could become activated signalling 'familiarity', and triggering subsequent stages of information retrieval and name retrieval. This framework inspired a computational model of face recognition (Burton, Bruce and Johnston's 1990 Interactive Activation and Competition (IAC) model) which has been augmented over the years to incorporate name recognition (Valentine, Brédart, Lawson & Ward, 1991) and voice recognition (Ellis, Jones & Mosdell, 1997; Stevenage, Hugill & Lewis, 2012). In essence, the Bruce and Young framework, and all subsequent variants, share the capacity to articulate voice processing as a separate yet parallel pathway to face and name processing within an overarching model of person perception.

The separate yet parallel nature of face, name and voice processing provides explanation for a number of clinical conditions, each of which represents a deficit of one pathway in the absence of a deficit in the other pathways. *Prosopagnosia*, for example, represents an inability to recognise individuals from their face alone (see Barton, 2008) and is usually indicated by damage to either the right fusiform face area (FFA) or the right anterior temporal lobe (ATL). Whilst some patients demonstrate both face and voice deficits and the term has not always been used appropriately (Gainotti, 2010), there are cases that show a pure deficit in face processing, supporting the existence of a discrete face processing pathway. In a similar vein, *anomia* represents an inability to retrieve a target person's name despite being able to retrieve semantic information such as their occupation, and despite being able to retrieve names of other objects (Harris & Kay, 1995; Semenza & Zettin, 1988). It is usually, but not specifically, associated with damage to the left temporal pole. In such individuals, the difficulty with name retrieval cannot merely be attributed to the uniqueness of the name, as particular celebrity catchphrases may be remembered (Harris & Kay, 1995). This supports the notion of a discrete name processing pathway. Finally, and relatively recently, cases have been identified who have an inability to recognise individuals from their voices whilst having a spared capacity to process faces and names (Garrido *et al.*, 2009; Hailstone *et al.*, 2010). The term *phonagnosia* has been applied to these individuals, with damage indicated in the right anterior temporal lobe (ATL) or the right superior temporal gyrus regions. The existence of these patients supports the notion of a discrete voice processing pathway. This hemispheric specialisation of face and voice processing within the right hemisphere, and name processing within the left hemisphere is mirrored from the patient population to the non-clinical population (Gainotti, 2011). Moreover, the fact that this pattern is shown for familiar and for unfamiliar targets suggests that these lateralisation effects exist for both perceptual and representational levels of processing (Gainotti, 2013).

The Integration of Faces and Voices for Identification

Interestingly, Hoover, Demonet and Steeves (2010) have recently described the case of SB, a patient with acquired prosopagnosia who compensated for his inability to process faces

by shifting the traditional balance of face and voice processing to emphasize the voice recognition pathway. This makes explicit the suggestion that whilst face and voice pathways are anatomically and neurologically distinct, they may functionally interact. Several instances exist within the literature to support the suggestion of an interaction between face and voice processing pathways. For example, Sheffert and Olson (2004) noted the benefit gained when learning voices if the voice was presented alongside the face at study compared to the presentation of the voice alone. Similarly, Armstrong and McKelvie (1996) and Legge, Grossman & Pieper (1984) demonstrated a benefit when recognising newly learned voices if the face accompanied the voice at test than if the voice was presented alone. These effects may owe something to the demonstration of an above-chance capacity to pair the voice and face of an unfamiliar target (see Kamachi, Hill, Lander & Vatikiotis-Bateson, 2003; Mavica & Barenholtz, 2013). Whilst Mavica and Barenholtz (2013) were unable to demonstrate a basis for this matching capacity, the demonstrable ability to link faces and voices may be important in learning and recognition studies.

Arguably one of the strongest pieces of evidence for an interaction between face and voice processing is the demonstration of an identity after-effect provided by Zäske, Schweinberger and Kawahara (2010). In an elegant study, adaptation was provided to a personally familiar voice (A) prior to the presentation of a second morphed voice (created to represent an ambiguous identity along a continuum between A and B). The results revealed a significant bias to identify the second target as more similar to B following adaptation to A (the after-effect; see also Latinus & Belin, 2012). Importantly for the current discussion, these vocal after-effects were noted following adaptation to A's voice (Experiment 1) and to A's face (Experiment 2). This provides a powerful demonstration of the capacity for face and voice processing to interact.

Two further demonstrations of this interaction are relevant. Both draw on the priming methodology in which prior presentation of one stimulus facilitates or primes the subsequent response to a second stimulus. In this regard, Stevenage, Hugill and Lewis (2012) explored identity priming in which a stimulus is processed more quickly at time 2 given its prior presentation at time 1. Within-modality identity priming was readily observed and was expected given an extensive priming literature (see Bruce, Carson, Burton & Kelly, 1998; Schweinberger, Herholz & Stief, 1997). Specifically, faces primed the processing of subsequent faces, and voices primed the processing of subsequent voices. However, the additional demonstration of cross-modality priming effects in this study confirmed an interaction between the face and voice processing pathways (see also Schweinberger, Robertson & Kaufmann, 2007), and was particularly strong when faces primed subsequent voice processing (Stevenage, Hugill & Lewis, 2012).

In a second study, Stevenage Hale, Morgan & Neil (2012) used an associative priming methodology in which the prior presentation of one stimulus facilitates the subsequent processing of a second, semantically- related stimulus. For example, the presentation of comedian Stan Laurel facilitates the subsequent processing of his comedy partner Oliver Hardy. Whilst within-modality associative priming is well documented (the face of one person primes the face of their associate; Bruce & Valentine, 1986), cross-modality associative priming has only recently been demonstrated (the name of one person can prime the face of their associate; Schweinberger, 1996; Wiese & Schweinberger, 2008). Stevenage *et al.* (2014) extended this paradigm to show cross-modality associative priming between faces and voices when sufficient time was allowed for the processing of a voice at the prime stage. Taken together, the learning effects, recognition effects, after-effects, and priming effects here all suggest that, whilst face and voice processing may proceed

independently, they have functional interactions which can be revealed through a number of behavioural tasks.

The literature on integration between face and voice goes beyond mere co-presentation. There is also an interesting literature on the temporal synchrony required between the animated face and the voice. Notably for example, Calvert, Brammer and Iverson (1998) present the synchrony of speaking faces as one of the most vivid examples of audiovisual integration. Benefits include the capacity to use congruous visual speech signals such as lip reading to help interpret speech amidst noise (see Sumby & Pollack, 1954). Additionally, compelling illusions exist when the face and the voice are incongruous (McGurk & MacDonald, 1976). With this in mind, a series of innovative studies by Schweinberger and colleagues have identified the critical temporal window within which co-presentation of the face and voice can facilitate identification tasks. Using both behavioural measures (Robertson & Schweinberger, 2010) and ERP measures (Schweinberger, Kloth & Robertson, 2011), the benefits of audiovisual integration are only felt when the voice is presented within the bounds of a 100ms auditory lead and a 300ms auditory lag relative to the face.

A Relative Weakness of Voice Recognition

Against this backdrop of examples illustrating a benefit when face and voice are presented together, there are, however, a number of examples which demonstrate significant interference when the face and voice are co-presented. Collectively, these suggest that, whilst the face and voice will often combine to mutually reinforce processing, the face will nevertheless dominate if the face and voice are incongruous.

Perhaps the best known example of this is the Facial Overshadowing effect (Cook & Wilding, 1997). This describes the finding that, when participants attempt to recognise a once-heard voice, they perform better when that voice has been presented in isolation at study rather than when presented alongside its face (Cook & Wilding, 1997; McAllister, Dale, Bregman, McCabe & Cotton, 1993). Stevenage, Howland and Tippelt (2011) extended this design to show that this audiovisual impairment was only demonstrated during voice recognition; in comparison, face recognition remained strong and stable no matter whether participants studied an isolated face or an audiovisual combination. These data suggest that the co-presentation of a face has the capacity to distract from the task of voice recognition, causing a measurable deficit in performance.

Whilst this might be expected when target individuals are previously unfamiliar, it is striking to see a similar facial overshadowing effect when targets are highly familiar celebrities (Stevenage, Neil & Hamlin, 2014). In an intricate design, face recognition was examined when faces were presented in the company of the relevant voice (identical condition), a semantically associated celebrity voice (related condition) and an irrelevant celebrity voice (unrelated condition). Face recognition remained strong and stable no matter what the identity of the accompanying voice (Experiment 1). In contrast, when voice recognition was examined in the presence of either the relevant face, a semantically related face or an irrelevant face, voice recognition was substantially and significantly affected by the identity of accompanying face: Performance was good when face and voice depicted the same individual (identical condition), but was significantly impaired to a predictable degree when the face and voice were merely related or unrelated.

In a similar vein, Hughes and Nicholson (2010) explored participants' capacity for self-recognition when presented with a face only, a voice only, or an audiovisual combination. Their results revealed superior performance when presented with the face than when presented with the voice, suggesting the primacy of the visual modality, However, again the results showed a deficit in the

audiovisual condition compared to the two single-modality conditions. What is striking about these results is that this facial overshadowing effect occurred despite the use of overlearned stimuli such as one's own voice.

Voices as a Weak Cue to Identity

This exploration of voice processing in the context of face processing may, on reflection, be an unhelpful way of examining voice recognition capacity. After all, when the face is present, we may have very little need to attend to the voice as a cue to identity. Consequently, it may only be possible to determine our capability with voice processing when performance is assessed in isolation of faces. A considerable literature now exists in this vein.

Surprisingly, when the competing demands of the face are separated out, the voice remains relatively weak as a cue to identity. A series of methodologies combine to support this view. For example, voices are less well recognised, and elicit significantly more 'familiar only' experiences compared to faces (Ellis, Jones & Mosdell, 1997; Hanley, Smith & Hadfield, 1998) and, in fact, recognition performance from faces and voices can only be equated when the faces are substantially blurred (Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009). In addition, when care is taken to balance recognition levels through blurring the faces, the capacity to retrieve semantic information about an individual remains substantially weaker when presented with the voice rather than the face. Notably, this remains the case when trying to retrieve semantic details about celebrities (Hanley & Damjanovic, 2009; Hanley, Smith & Hadfield, 1998), personally familiar individuals (Barsics & Brédart, 2011; Brédart, Barsics & Hanley, 2009) or newly learned individuals (Barsics & Brédart, 2012a).

Finally, the voice is shown to be a weaker cue relative to the face when participants try to recall episodic information about individuals. Using a modified version of the Remember/Know paradigm (Gardiner & Richarson-Klavehn, 2000), participants indicated their ability to retrieve a specific instance or episode in which they encountered a target individual. Capacity to recollect such an instance is reported as a 'remember' state and is used as evidence of episodic retrieval, whereas a general familiarity in the absence of the recollection of an episode is reported as a 'know' state and indicates a lack of episodic retrieval. Across a series of studies using voices and blurred faces as the stimuli, the voice emerges as a weaker cue to episodic retrieval. It elicits fewer 'remember' states, and more 'know' states compared to the face both when responding to celebrity targets (Damjanovic & Hanley, 2007) and to targets that are personally familiar (Barsics & Brédart, 2011).

These data as a whole are compelling in demonstrating that, despite balancing the recognisability of the input face and voice, the voice is relatively poor as a cue to identity. With this in mind, it is reasonable to assume that voices would also be relatively vulnerable, compared to faces, to any factors that may disrupt processing.

Vulnerability of Voices

One study designed to test the vulnerability of voice processing is provided by Stevenage, Neil, Barlow, Dyson, Eaton-Brown and Parsons (2013). They explored the capacity of participants to withstand the effects of interference. In a standard sequential same/different matching task, participants experienced interference through the presentation of either 0, 2, or 4 distractors within a fixed duration interval in between study and test. With face recognition interrupted by face distractors, and voice recognition interrupted by voice distractors, the results supported all predictions. Face recognition remained high no matter what level of distraction was provided. In contrast, voice recognition was significantly impaired as soon as any distraction was introduced. Moreover, this pattern of performance remained evident regardless of whether distractors were similar to the target

(same sex – Experiment 1) or not (different sex – Experiment 2), suggesting a significant and specific impairment in being able to create and retain a strong mental representation of a target voice.

A second series of studies can also be taken to demonstrate the vulnerability of voice recognition when processing is disrupted. In this series, voice recognition was made difficult through the manipulation of the language being spoken. As above, results supported all predictions - voice recognition was significantly impaired when the target was speaking in an unfamiliar language compared to when speaking in a native tongue. In this regard, Myers (2001) notes the 'all sound alike' phenomenon in which speakers with unfamiliar accents can be difficult to distinguish. This has become known as the 'other accent effect' and perhaps shares a common basis with the 'other race effect' within the face recognition literature. In both cases, a lack of expertise with a particular subset of stimuli, such as non-native speakers or other-race faces, can result either in a reduced sensitivity when trying to discriminate between them, or a reliance on inappropriate characteristics more suited to the more-often experienced stimuli (see Valentine, 1991). The result is a comparatively poor level of processing for the minority set.

In the voice recognition field, several studies demonstrate this point. For example, Goggin, Thompson, Strube and Simental (1991) reported an elegant series of experiments in which monolingual listeners identified bilingual speakers with significantly greater accuracy and confidence when they spoke in a familiar language as opposed to an unfamiliar language (Experiments 1 and 2). Additionally, monolingual listeners showed an own-language bias when recognising bilingual speakers, but bilingual listeners did not (Experiment 3). As a whole, these studies provide compelling evidence to suggest that when speaker and listener variables are respectively held constant, other-accent effects emerge.

Phillipon, Cherryman, Bull and Vrij (2007) provide support for these conclusions. They asked English listeners to identify either English speakers or French speakers within a lineup paradigm. Results suggested that performance was significantly better when the target spoke in a familiar (English) language rather than in an unfamiliar (French) language. More specifically, there was a poor rate of correct identifications and a high rate of false alarms when the targets spoke in the unfamiliar language. These results are notable because of the design that Phillipon *et al.* (ibid) used. In particular, their use of long clips both at study (45–50 seconds) and at test (25–34 seconds), and their use of a methodology in which participants could listen to targets twice prior to response, may have been expected to minimise other-accent effects. Their emergence despite these methodological considerations confirms the vulnerability of voice recognition.

The final study in this sequence is presented by Stevenage, Clarke and McNeill (2012). Rather than varying the language in which the target spoke, Stevenage *et al.* varied the accent that they spoke with. Thus, all speakers spoke in English, yet some had a distinct Glasgow accent, and some had a distinct Southern English accent. English and Scottish listeners then took part in a lineup study with English and Scottish accented speakers. Consequently, the experimental design enabled comprehension of all speakers to be held constant, but familiarity with the speaker accent to be varied. The results accorded with those studies reviewed previously in that performance in the lineup task was significantly better when the target spoke in a familiar rather than an unfamiliar accent. In particular, there was a greater false alarm rate in target-present lineups, and fewer correct rejections and more false alarms in target-absent lineups, when the speaker had an unfamiliar accent.

Interestingly within this study, an asymmetry was noted in the other-accent effect, with English listeners affected more by the Scottish accent than Scottish listeners were

affected by the English accent. In common with a similar asymmetry in the other-race effect with faces, this finding may usefully be set in the context of differential expertise effects. Indeed, in both the other race effect and the other accent effect, one participant group invariably has more exposure to their unfamiliar stimulus group than the other participant group because those stimuli are more commonly experienced within their environment. The fact that both face processing and voice processing show a clear parallel in this regard provides a nice link between the cognitive tasks underpinning face and voice processing. The real value, however, of the current study is that, unlike previous studies, it separates out the issue of comprehensibility from the issue of accent in that the content of both familiar and unfamiliar accented speakers was comprehensible. The fact that the unfamiliar accent still distracted from the capacity to identify the target speaker both supports and confirms the vulnerability of voice recognition to circumstances which increase task difficulty.

Not All Voices are Created Equal: Robust Exemplars

Up to this point, the literature has suggested a separate pathway for voice recognition compared to face recognition, but a relative weakness of the voice recognition pathway. Consequently, there is a dominance of the face in situations where both are present, or there is a mismatch. It is possible to explain the relative weakness of voice recognition compared to face recognition in terms of either a differential utilisation account or a differential sensitivity account (see Stevenage, Hugill & Lewis, 2012). According to the differential utilisation account, voices are simply used less often as a cue to identity because they are so often accompanied by the face. In terms of the differential sensitivity account, listeners are less able to discriminate between vocal characteristics than viewers are to discriminate between facial characteristics. The result from either perspective is that voice recognition is poorer than face recognition.

As a consequence, it is tempting to suppose that a relatively weak process will be relatively more vulnerable when difficulty arises, and a body of literature has been brought to bear to support that supposition. However, notable exceptions do exist, and it is worth focussing on those here in terms of the refinements to thinking that they encourage.

Perhaps of most interest in this regard is the observation of what happens when voices are played backwards (van Lanker, Kreiman & Emmorey, 1985; see also Goggin *et al.*, 1991, Experiment 4). Participants in this study were asked to listen to 45 celebrity voices. Performance was striking in that participants were able to recognise nearly 27 per cent of targets in an old/new task, and nearly 70 per cent of targets in a 6-alternate forced-choice (6AFC) task. Remarkably, participants remained able to recognise over 57 per cent of targets in the 6AFC task when those targets were played backwards. What is notable for the present discussion is that performance on these backwards voices varied substantially across the targets, with some targets being as recognisable when played backwards as when played forwards. This raises the likelihood that averaged indicators of voice recognition performance hide quite substantial inter-voice (or inter-item) differences.

To us, this discussion fits well with the literature on distinctiveness, and links with the recent work of Barsics and Brédart (2012b, see also Mullenix, Ross, Smith, Kuykendall, Conard & Barb, 2011; Sorensen, 2012; Zetterholm, Sarwar, Thorvaldsson & Allwood, 2012 for further examples of vocal distinctiveness effects). Within their study, Barsics and Brédart explored the retrieval of episodic and semantic information from distinctive and typical faces and voices. In a very well-grounded experiment which echoed work exploring the distinctiveness advantage in face recognition, the results supported a role for vocal distinctiveness in voice processing. Specifically, a distinctiveness-advantage

emerged when retrieving semantic information from a voice cue. Interestingly though, distinctive voices still elicited less semantic retrieval than typical faces suggesting again the dominance of the facial pathway over the vocal pathway. Nevertheless, their results converge with those of van Lanker *et al.* in highlighting item effects within voice recognition.

With this in mind, Neil and colleagues conducted a series of studies to explore the performance benefits that may be demonstrated for *strong* voices (Neil, Stevenage & Parsons, in prep). Voices were defined as strong either through natural means (they were rated as distinctive rather than typical) or through artificial means (they were heard five times rather than heard once). Our hypothesis was that *strong* voices would behave rather like van Lanker *et al.'s* distinctive voices and would elicit performance benefits in a recognition task.

Using the sequential same/different matching task described earlier, we asked participants to study and then to recognise a series of *strong and weak* voices. Interference was provided such that participants heard either 0 distractors or 4 distractors during the gap between study and test. Our prediction was that accuracy on the same/different matching task would be compromised by the introduction of interference, but that *strong* voices, whether through natural or artificial means, would be affected less by interference than *weak* voices. The results supported all predictions: distinctive voices (Experiment 1) and repeated voices (Experiment 2) each showed interference effects but these were significantly reduced in magnitude compared to the interference effects for the corresponding *weak* voices. Despite this, performance in each condition was above chance, confirming that these results could not be attributed to floor effects when recognising *weak* stimuli.

Together, the results of van Lanker *et al.* (1985) and of Neil *et al.* (above) suggest that item effects exist and may best be understood in the context of a substantial literature on distinctiveness effects in recognition tasks. In line with distinctiveness effects in the face recognition literature, it is useful to align the current results with a multidimensional space framework in which items that stand out on one or more dimension consequently stand out within the multidimensional space. Confusion with similar others, or' near neighbours', is consequently less likely and the item is better recognised as a result. The work of Baumann and Belin (2010) highlights the importance of the fundamental frequency (pitch) of a voice along with variation in formant information, and this is useful in defining what these dimensions may be within a vocal similarity space.

The Influence of Vocal Emotion on Voice Processing

Of course, it need not be the case that a voice is distinctive because of some set of stable and intransient characteristics. Indeed, a voice may attain distinctiveness on a particular occasion through factors that momentarily make it stand out, such as through heightened emotion. Emotion is perhaps of particular interest because a literature exists to suggest that emotional stimuli may attract and hold attention better than neutral stimuli. For instance, emotional faces hold attention better than neutral stimuli amongst anxious individuals in a dot-probe task (see for example Fox, Russo, & Dutton, 2002) and for all individuals in an attentional blink paradigm (Milders, Sahraie, Logan & Donnellon, 2006; de Jong & Martens, 2007). Given the parallels between face processing and voice processing, the processing of emotional voices represents a natural next question.

In this regard, one might hold the view that emotional voices may also be *strong* voices, capable of attracting and holding attention, and supporting improved performance in a subsequent recognition context. A series of studies exist to allow a test of this hypothesis. However, surprisingly, the results tend to run counter to expectation. For example, Saslove and Yarmey (1980) conducted a

voice line-up task. At the initial study phase, participants overheard an angry speaker for 11 seconds. At test, they heard the target again, plus four distractor voices repeating the original message, with the target either speaking in the original angry tone, or in a neutral conversational tone. The results suggested that performance was significantly worse when the speaker's tone of voice had changed between study and test. In fact, performance was no better than chance in that condition. Consequently, the presentation of an emotional voice at study did not facilitate subsequent voice recognition wholesale.

Similarly, Read and Craik (1995) explored the impact of emotional tone on voice recognition via a voice lineup task. In their study, participants first heard a series of six speakers clips which were rated for emotional intensity. The target clip ('Help me, help me, Oh God, help me!') was significantly more emotional than the other five clips. After a break of 17 days, participants were asked to identify the speaker who had uttered the emotional clip from amongst a set of 6 speakers. Using a six-alternate forced choice task, with two opportunities to hear each speaker, participants showed a 66% recognition rate when emotional tone of voice was unchanged between study and test. In contrast, performance fell to 20% and was no better than chance when the emotional tone of voice changed from angry to a neutral conversational tone, despite a lengthy 20 second clip at this latter test opportunity. This result emerged whether unfamiliar voices (Experiment 1) or moderately familiar voices (Experiment 2) were used as stimuli.

The results of both studies suggest that rather than improving recognition levels, emotion may impair them, especially when the emotional tone changes between study and test. However, both studies lack the data from a control condition in which the target spoke with a neutral tone at study and at test. More recent work by Ohman, Eriksson and Granhag (2013) provides this neutral baseline, and casts some doubt on previous results, revealing no effects of emotional change between study and lineup. Whilst, floor effects may have precluded any impact of emotion in this study, the fact that the emotional voice again did not facilitate subsequent recognition relative to the neutral base remains a surprise.

In a final set of studies, Stevenage, Neil, Hearsum & Long (in prep) have investigated the role of emotional tone using a same/different matching task rather than a lineup task. This enabled greater statistical power through the examination of performance across a number of trials, rather than in a single lineup trial. In Experiment 1, the stimuli were carefully designed such that targets uttered a constant phrase in either an angry, happy or neutral tone of voice. Consequently, semantic content and linguistic variability were controlled. At test, all speakers uttered a second phrase in a neutral tone of voice, allowing the impact of emotional tone, and emotional change to be assessed. The results were clear and confirmed that, relative to a neutral study and test voice, an emotional voice at study made subsequent recognition worse rather than better at test. Moreover, the results of Experiment 2 perhaps provide the cleanest test of this issue through using a fully crossed design. Specifically, participants either heard neutral voices at study and test (NN), emotional voices at study and test (EE), neutral voices at study with emotional voices at test (NE) or emotional voices at study with neutral voices at test (EN). The results of this study provide the strongest indication yet of a decline in performance when emotional tone changed, irrespective of whether the emotional tone was presented at study (EN) or at test (NE). Consequently, these data confirmed the findings of Saslove and Yarmey (1980) and Read and Craik (1995) and suggested that emotional voices did not produce the levels of performance associated with 'strong' voices, especially when emotional tone changed from one instance to the next.

In the context of the preceding discussion around distinctiveness effects, and a benefit

gained from the presentation of *strong* voices, the results of these emotional voices studies are surprising. One might have anticipated that emotional voices would attract attention and thus, that performance for these voices, whether emotional or neutral at test, would echo the advantage shown for other *strong* voices. The results are entirely contrary to this expectation and force a reconsideration of the framework within which we have been considering voice recognition. In this vein, it is useful to consider identification alongside other tasks for which the voice may be of valuable. Belin, Bestelmeyer, Latinus and Watson's (2011) concept of the voice as an 'auditory face' is particularly appealing in this regard, and is used as a framework to help shape the rest of this discussion.

A Broader Perspective

To this point, we have reviewed a host of empirical studies of voice recognition. They collectively support the view that voice recognition proceeds through a separate yet parallel pathway to face (and name) recognition. The voice recognition pathway, however, emerges as a relatively weak route to person identification as noted through competition effects and facial overshadowing effects. Moreover, even when the voice is presented alone as a cue to identity, it is less successful in eliciting person-related details such as semantic, episodic or name information compared to the face. Nevertheless, not all voices are created equal in that some voices stand out from the crowd and can be recognised well despite quite substantial manipulation such as a backwards audio track. In light of this, it is surprising to find that emotional voices are not amongst those that are processed well.

The work of Belin and colleagues is important in this regard because it presents the first consideration of voice recognition in the context of other tasks for which the voice may be important. Notably, Belin, Bestelmeyer, Latinus and Watson (2011) describe the voice as an 'auditory face' from which we are able to process three things: vocal content (what does the speaker want?), vocal affect (how does the speaker feel?) and vocal identity (who is the speaker?).

This view of the voice as an auditory face led Belin *et al.* to propose a valuable heuristic for voice processing as a whole. Specifically, they suggested that, following some common stage of 'structural encoding', speech, identity and affect are extracted through separate pathways, each drawing on different neural structures. A double dissociation between the deficits shown by aphasic patients (inability to process vocal content) and phonagnosic patients (inability to process vocal identity) lends support to Belin *et al.*'s articulation of separate speech and identity pathways within voice processing. Similarly, the fact that phonagnosic patients tend to be able to process vocal affect (see Garrido *et al.* 2009; Hailstone *et al.* 2010) lends support to the articulation of separate affect and identity pathways within voice processing. As a whole, this framework provides value because it sets the task of voice recognition within a broader ecological and psychosocial context in which the voice represents the 'whole person' (see Sidtis & Kreiman, 2012).

Crucially, however, evidence exists to support the view that these separate pathways interact with one another during normal voice processing, and are better considered as partially dissociated pathways. For example von Kriegstein, Smith, Patterson, Kiebel and Griffiths (2010) provide fMRI evidence to suggest that speaker characteristics modulate neural activity in speech processing areas. Similarly, speaker familiarity affects performance in basic linguistic tasks (Nygaard & Pisoni, 1998) and can help with the processing of semantic content (Creel & Tumlin, 2011). Moreover, the three voice processing pathways are suggested to interact with corresponding face processing pathways, supporting the demonstrations previously noted to occur during audio-visual integration (see Campanella & Belin, 2007).

What is exciting for the present discussion is the addition of a view expressed by Goggin

et al. (1991). They suggested that, whilst we have the capacity to process all three aspects of vocal information, we cannot help but prioritise some aspects over others. Specifically, Goggin *et al.* suggested that we may automatically orient our attention to process speech content and, at times, this may be to the detriment of other aspects of vocal processing. Our suggestion here is that Belin *et al.*'s concept of partially dissociated voice processing pathways can be combined with Goggin *et al.*'s concept of prioritisation of pathways, to provide a novel account of the collective findings reviewed within this paper. It then becomes clear that when attentional resources are required to process speech content (because language or accent is unfamiliar) then speaker recognition is impaired. Similarly, the previously surprising results from studies of emotional voice recognition become understandable: When attentional resources are required to process affect (because affect is heightened for example) then speaker recognition is again impaired. Given the primacy that Goggin *et al.* ascribe to the processing of content over identity, it remains to be seen whether speaker recognition may ever dominate over speech or affect processing. However, all other things being equal, the distinctiveness effects noted earlier may provide tentative support for such a suggestion.

Concluding Remarks

Within the present paper, we have provided a review of the growing literature on voice recognition. Taking a lead from the face recognition literature, a substantial body of work now exists which highlights notable parallels between the two recognition processes. Whilst surface characteristics differ substantially between the face and voice, the underlying processes share considerable overlap, and have been considered as parallel pathways within the person recognition system. What has become clear, however, is that the voice recognition pathway is weaker than the face recognition pathway, showing overshadowing and interference effects which diminish the capacity to recognise the voice. Certainly, these effects would suggest great caution when evaluating the credibility of an earwitness within a court of law.

The capacity of the face and voice recognition pathways to interact within a multimodal person recognition system now has a substantial level of support within the literature. However, we discuss here the value of an even larger integrative framework in which recognition from face or voice is set alongside the processing of facial and vocal affect, and vocal speech. The capacity of these broader pathways to interact, and to compete when processing demands are great, offers a novel and exciting framework. We suggest that it may provide a parsimonious explanation for a number of different findings, and that it may act as a powerful source of predictions for face and voice researchers moving forwards.

Acknowledgements

This work was supported by EPSRC Grant (EP/J004995/1 SID: An Exploration of SuperIdentity). Colleagues on this grant are thanked for their helpful contributions to the current work.

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: http://dx.doi.org/10.1080/00221309.1996.9921278

Barenholtz, E. (2013). Matching Voice and Face Identity From Static Images. *Journal of Experimental Psychology: Human Perception and Performance, 39(2)*, 307–312. DOI: http://dx.doi.org/10.1037/a0030945

Barsics, C., & **Brédart, S.** (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition, 20*, 303–308. DOI: http://dx.doi.org/10.1016/j.concog.2010.03.008

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

527–534. DOI: http://dx.doi.org/1080/09658211.2012.683012

Barsics, C., & **Brédart, S.** (2012b). Access to semantic and episodic information from faces and voices: Does distinctiveness matter? *Journal of Cognitive Psychology, iFirst* 1–7. DOI: http://dx.doi.org/10.1080/20445911.2012.692672

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: Lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology, 2*, 197–225. DOI: http://dx.doi.org/10.1348/174866407X214172

Baumann, O., & **Belin, P.** (2010). Perceptual Scaling of Voice Identity: Common Dimensions for Different Vowels and Speakers. *Psychological Research, 74*, 110–120. DOI: http://dx.doi.org/10.1007/s00426-008-0815-z

Beauchemin, M., et al. (2011). Mother and Stranger: An Electrophysiological study of Voice Processing in Newborns. *Cerebral Cortex, 21(8)*, 1705–1711. DOI: http://dx.doi.org/10.1093/cercor/bhq242

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

Brédart, S., & **Barsics, C.** (2012). Recalling semantic and episodic information from faces and voices: A face advantage. *Current Directions in Psychological Science, 21(6)*, 378–381. DOI: http://dx.doi.org/10.1177/0963721412454876

Brédart, S., Barsics, C., & **Hanley, R.** (2009). Recalling semantic information about personally known faces and voices. *European Journal of cognitive Psychology, 21(7)*, 1013–1021. DOI: http://dx.doi.org/10.1080/09541440802591821

Bruce, V., Carson, D., Burton, A. M., & **Kelly, S. W.** (1998). Prime time advertisements: Repetition Priming from Faces seen on Subject Recruitment Posters. *Memory and Cognition, 26(3)*, 502–515. DOI: http://dx.doi.org/10.3758/BF03201159

Bruce, V., & **Young, A.** (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305–327. DOI: http://dx.doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., Bruce, V., & **Johnston, R. A.** (1990). Understanding face recognition with an Interactive Activation and Competition model. *British Journal of Psychology, 81*, 361–380. DOI: http://dx.doi.org/10.1111/j.2044-8295.1990.tb02367.x

Calvert, G. A., Brammer, M. J., & **Iverson, S. D.** (1998). Crossmodal Identification. *Trends in Cognitive Science, 2*, 247–253. DOI: http://dx.doi.org/10.1016/S1364-6613(98)01189-9

Campanella, S., & **Belin, P.** (2007). Integrating Face and Voice in Person Perception. *Trends in Cognitive Science, 11(12)*, 535–543. DOI: http://dx.doi.org/10.1016/j.tics.2007.10.001

Cook, S., & **Wilding, J.** (1997). Earwitness Testimony 1: Voices, Faces and Context. *Applied Cognitive Psychology, 11*, 527–541. DOI: http://dx.doi.org/10.1002/(SICI)1099-0720(199712)11:6<527::AID-ACP483>3.0.CO;2-B

Creel, S. C., & **Tumlin, M. A.** (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65(3)*, 264–285. DOI: http://dx.doi.org/10.1016/j.jml.2011.06.005

Damjanovic, L., & **Hanley, J. R.** (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition, 35(6)*, 1205–1210. DOI: http://dx.doi.org/10.3758/BF03193594

De Jong, P., & **Martens, S.** (2007). Detection of Emotional Expressions in Rapidly Changing Facial Displays in High- and Low-Socially Anxious Women. *Behavior Research and Therapy, 45*, 1285–1294. DOI: http://dx.doi.org/10.1016/j.brat.2006.10.003

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. DOI: http://

dx.doi.org/10.1111/j.2044-8295.1997.tb02625.x

Fox, E., Russo, R., & **Dutton, K.** (2002). Attention Bias for Threat: Evidence for Delayed Disengagement from Emotional Faces. *Cognition and Emotion, 16*, 355–379. DOI: http://dx.doi.org/10.1080/02699930143000527

Gainotti, G. (2010). Not all patients labelled as prosopagnosia have a real prosopagnosia. *Journal of Clinical and Experimental Neuropsychology, 32(7)*, 763–766. DOI: http://dx.doi.org/10.1080/13803390903512686

Gainotti, G. (2011). What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia, 49(9)*, 2273–2282. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2011.04.027

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

Gardiner, J. M., & **Richardson-Klavehn, A.** (2000). Remembering and Knowing. In E. Tulving and F.I.M. Craik (Eds.) *Handbook of Memory* (pp 229–244). Oxford: Oxford University Press.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., Schweinberger, S. R., Warren, J. D., & **Duchaine, B.** (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47(1)*, 123–131. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2008.08.003

Goggin, J. P., Thompson, C. P., Strube, G., & **Simental, L. R.** (1991). The Role of Language Familiarity in Voice Identification. *Memory and Cognition, 19*, 448–458. DOI: http://dx.doi.org/10.3758/BF03199567

Hailstone, J. C., Crutch, S. J., Bestergaard, M. D., Patterson, R. D., & **Warren, J. D.** (2010). Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia, 48*, 1104–1114. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2009.12.011

Hanley, J. R. (2014). Accessing stored knowledge of familiar people from faces, names and voices: A review. *Frontiers in Bioscience, E6*, 198–207. DOI: http://dx.doi.org/10.2741/E702

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. DOI: http://dx.doi.org/10.1080/09658210903264175

Hanley, J. R., Smith, S. T., & **Hadfield, J.** (1998). I recognize 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*. DOI: http://dx.doi.org/10.1080/713755751

Harris, D. M., & **Kay, J.** (1995). I recognize your face but I can't remember your name: Is it because names are unique? *British Journal of Psychology, 83*, 45–60.

Hoover, A. E. N., Demonet, J. F., & **Steeves, J. K. E.** (2010). Superior voice recognition in a patient with acquired prosopagnosia and object agnosia. *Neuropsychologia, 48(13)*, 3725–3732. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2010.09.008

Hughes, S. M., & **Nicholson, S. E.** (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness and Cognition, 19(4)*, 1124–1134. DOI: http://dx.doi.org/10.1016/j.concog.2010.03.001

Kamachi, M., Hill, H., Lander, K., & **Vatikiotis-Bateson, E.** (2003). Putting the face to the voice: Matching Identity across Modality. *Current Biology, 13*, 1709–1714. DOI: http://dx.doi.org/10.1016/j-cub.2003.09.005

Latinus, M., & **Belin, P.** (2011). Human Voice Perception. *Current Biology, 21(4)*,

R143-R145. DOI: http://dx.doi.org/10.1016/j.cub.2010.12.033

Latinus, M., & **Belin, P.** (2012). Perceptual Auditory Aftereffects on Voice Identity using Brief Vowel Stimuli. *PLoS ONE, 7(7)*, e41384. DOI: http://dx.doi.org/10.1371/journal.pone.0041384

Legge, G. E., Grossman, C., & **Pieper, C. M.** (1984). Learning Unfamiliar Voices. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 298–303. DOI: http://dx.doi.org/10.1037/0278-7393.10.2.298

Mavica, L. W., & **Barenholtz, E.** (2013). Matching Voice and Face Identity from Static Images. *Journal of Experimental Psychology: Human Perception and Performance, 39(2)*, 307–312. DOI: http://dx.doi.org/10.1037/a0030945

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 Identification. *Basic and Applied Social Psychology, 14(2)*, 161–170. DOI: http://dx.doi.org/10.1207/s15324834basp1402_3

McGurk, H., & **MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748. DOI: http://dx.doi.org/10.1038/264746a0

Milders, M., Sahraie, A., Logan, S., & **Donnellon, N.** (2006). Awareness of Faces is Modulated by their Emotional meaning. *Emotion, 6*, 10–17. DOI: http://dx.doi.org/10.1037/1528-3542.6.1.10

Mullenix, J. W., Ross, A., Smith, C., Kuykendall, D., Conard, J., & **Barb, S.** (2011). Typicality Effects on Memory for Voice: Implications for Earwitness Testimony. *Applied Cognitive Psychology, 25(1)*, 29–34. DOI: http://dx.doi.org/10.1002/acp.1635

Myers, D. G. (2001). *Psychology*. New York: Worth Publishers.

Neil, G. J., Stevenage, S. V., & **Parsons, B.** (in prep). Protecting Voice Recognition from the Effects of Interference. (unpublished manuscript).

Nygaard, I. C., & **Pisoni, D. B.** (1998). Talker-Specific Learning in Speech Perception. *Perception and Psychophysics, 60*, 355–376. DOI: http://dx.doi.org/10.3758/BF03206860

Ohman, L., Eriksson, A., & **Granhag, P. A.** (2013). Angry Voices from the Past and Present: Effects on Adults' and Children's Earwitness Memory. *Journal of Investigative Psychology and Offender Profiling, 10(1)*, 57–70. DOI: http://dx.doi.org/10.1002/jip.1381

Phillipon, A. C., Cherryman, J., Bull, R., & **Vrij, A.** (2007). Earwitness Identification Performance: The Effect of Language, Threat, Deliberate Strategies and Indirect Measures. *Applied Cognitive Psychology, 21*, 539–550. DOI: http://dx.doi.org/10.1002/acp.1296

Read, D., & **Craik, F. I. M.** (1995). Earwitness Identification: Some Influences on Voice Recognition. *Journal of Experimental Psychology: Applied, 1(1)*, 6–18. DOI: http://dx.doi.org/10.1037/1076-898X.1.1.6

Robertson, D. M. C., & **Schweinberger, S. R.** (2010). The role of audiovisual asynchrony in person recognition. *Quarterly Journal of Experimental Psychology, 63(1)*, 23–30. DOI: http://dx.doi.org/10.1080/17470210903144376

Saslove, H., & **Yarmey, A. D.** (1980). Long-Term Auditory Memory: Speaker Identification. *Journal of Applied Psychology, 65(1)*, 111–116. DOI: http://dx.doi.org/10.1037/0021-9010.65.1.111

Schweinberger, S. R. (1996). How Gorbachov primed Yeltsin: Analysis of Associative Priming in Person Recognition by means of Reaction Times and Event Related Brain Potentials *Journal of Experimental Psychology, 22*, 1383–1407.

Schweinberger, S. R., Herholtz, A., & **Stief, V.** (1997). Auditory Long-Term Memory: Repetition Priming of Voice Recognition. *Quarterly Journal of Experimental Psychology, 50A(3)*, 498–517. DOI: http://dx.doi.org/10.1080/713755724

Schweinberger, S. R., Kloth, N., & **Robertson, D. M. C.** (2011). Hearing facial identities: Brain correlates of face-voice integration in person identification. *Cortex, 47(9)*, 1026–1037. DOI: http://dx.doi.org/10.1016/j.cortex.2010.11.011

Schweinberger, S. R., Robertson, D., & **Kaufmann, J. M.** (2007). Hearing Facial Identities. *Quarterly Journal of Experimental Psychology, 60*, 1416–1456. DOI: http://dx.doi.org/10.1080/17470210601063589

Semenza, C., & **Zettin, M.** (1988). Generating proper names: A case of selective inability. *Cognitive Neuropsychology, 5*, 711–721. DOI: http://dx.doi.org/10.1080/02643298808253279

Sheffert, S. M., & **Olson, E.** (2004). Audiovisual speech facilitates voice learning *Perception and Psychophysics, 66(2)*, 352–362. DOI: http://dx.doi.org/10.3758/BF03194884

Sidtis, D., & **Kreiman, J.** (2012). In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative Psychological and Behavioral Science, 46(2)*, 146–159. DOI: http://dx.doi.org/10.1007/s12124-011-9177-4

Sorensen, M. H. (2012). Voice line-ups: speakers' F0 values influence the reliability of voice recognitions. *International Journal of Speech Language and the Law, 19(2)*, 145–158. DOI: http://dx.doi.org/10.1558/ijsll.v19i2.145

Stevenage, S. V., Clarke, G., & **McNeill, A.** (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology, 24(6)*, 647–653. DOI: http://dx.doi.org/10.1080/20445911.2012.675321

Stevenage, S. V., Hale, S., Morgan, Y., & **Neil, G. J.** (2014). Recognition by Association: Within- and Cross-Modality Associative Priming with Faces and Voices. *British Journal of Psychology, 105(1)*, 1–16. DOI: http://dx.doi.org/10.1111/bjop.12011

Stevenage, S. V., Howland, A., & **Tippelt, A.** (2011). Interference in Eyewitness and Earwitness Recognition. *Applied Cognitive Psychology, 25(1)*, 112–118. DOI: http://dx.doi.org/10.1002/acp.1649

Stevenage, S. V., Hugill, A. R., & **Lewis, H. G.** (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology, 24(4)*, 409–419. DOI: http://dx.doi.org/10.1080/20445911.2011.642859

Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & **Parsons, B.** (2013). The effect of distraction on face and voice recognition. *Psychological research, 77(2)*, 167–175. DOI: http://dx.doi.org/10.1007/s00426-012-0450-z

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

Stevenage, S. V., Neil, G. J., Hearsum, S., & **Long, R.** (in prep). The Impact of Emotion when Recognising Voices. *Unpublished manuscript.*

Sumby, W. H., & **Pollack, I.** (1954). Visual contribution to speech intelligibility in noise. *Journal of Acoustic Society of America, 26*, 212–215. DOI: http://dx.doi.org/10.1121/1.1907309

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion and Race in Face Recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161–204. DOI: http://dx.doi.org/10.1080/14640749108400966

Valentine, T., Brédart, S., Lawson, R., & **Ward, G.** (1991). What's in a name? Access to information from people's names. *European Journal of Cognitive Psychology, 3*, 147–176. DOI: http://dx.doi.org/10.1080/09541449108406224

Van Lanker, D., Kreiman, J., & **Emmorey, K.** (1985). Familiar voice recognition: Patterns and Parameters. Part I: Recognition of Backwards Voices. *Journal of Phonetics, 13*, 19–38.

Von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Kiebel, S. J., & **Griffiths, T. D.** (2010). How the Human Brain Recognizes Speech in the Context of Changing Speakers. *Journal of Neuroscience, 30(2)*, 629–638. DOI: http://dx.doi.org/10.1523/JNEUROSCI.2742-09.2010

Wiese, H., & **Schweinberger, S. R.** (2008). Event-related Potentials indicate Different Processing to mediate Categorical and Associative Priming in Person Recognition. *Memory and Cognition, 34*, 1246–1263. DOI: http://dx.doi.org/10.1037/a0012937

Yovel, G. & **Belin, P.** (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences, 17(6)*, 263–271. DOI: http://dx.doi.org/10.1016/j.tics.2013.04.004

Zäske, R., Schweinberger, S. R., & **Kawahara, H.** (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research, 268*, 38–45. DOI: http://dx.doi.org/10.1016/j.heares.2010.04.011

Zetterholm, E., Sarwar, F., Thorvaldsson, V., & **Allwood, C. M.** (2012). Earwitnesses: the effect of type of vocal differences on correct identification and confidence accuracy. *International Journal of Speech, Language and the Law, 19(2)*, 219–237. DOI: http://dx.doi.org/10.1558/ijsll.v19i2.219

How to cite this article: Stevenage, S. V. and 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

Submitted: 15 March 2014 **Accepted:** 28 April 2014 **Published:** 20 May 2014

]u[*Psychologica Belgica* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS