





## RESEARCH ARTICLE

WILEY

# Are super-face-recognisers also super-voice-recognisers? Evidence from cross-modal identification tasks

Ryan E. Jenkins<sup>1</sup>  | Stella Tsermentseli<sup>1</sup> | Claire P. Monks<sup>1</sup>  |  
David J. Robertson<sup>2</sup>  | Sarah V. Stevenage<sup>3</sup> | Ashley E. Symons<sup>3</sup> | Josh P. Davis<sup>1</sup> 

<sup>1</sup>School of Human Sciences, Institute for Lifecourse Development, University of Greenwich, London, UK

<sup>2</sup>School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

<sup>3</sup>Department of Psychology, University of Southampton, Southampton, UK

## Correspondence

Ryan E. Jenkins, School of Human Sciences, Institute for Lifecourse Development, University of Greenwich, London, SE10 9LS, UK.

Email: r.e.jenkins@greenwich.ac.uk

## Funding information

University of Greenwich, Grant/Award Number: Vice Chancellor's Ph.D Scholarship (2017/2018)

## Summary

Individual differences in face identification ability range from prosopagnosia to super-recognition. The current study examined whether face identification ability predicts voice identification ability (participants:  $N = 529$ ). Superior-face-identifiers (exceptional at face memory and matching), superior-face-recognisers (exceptional at face memory only), superior-face-matchers (exceptional face matchers only), and controls completed the Bangor Voice Matching Test, Glasgow Voice Memory Test, and a Famous Voice Recognition Test. Meeting predictions, those possessing exceptional face memory and matching skills outperformed typical-range face groups at voice memory and voice matching respectively. Proportionally more super-face-identifiers also achieved our super-voice-recogniser criteria on two or more tests. Underlying cross-modality (voices vs. faces) and cross-task (memory vs. perception) mechanisms may therefore drive superior performances. Dissociations between Glasgow Voice Memory Test voice and face recognition also suggest voice-specific effects to match those found with faces. These findings have applied implications for policing, particularly in cases when only suspect voice clips are available.

## KEYWORDS

face recognition, policing, super-recognition, voice matching, voice recognition

The ability to identify a face or voice is important in many social and security-based contexts (e.g., border control, policing, surveillance). Face identification research using recognition-memory (e.g., McCaffery et al., 2018; Robertson et al., 2016; Russell et al., 2009) and simultaneous matching paradigms (e.g., Bate et al., 2019; Bate & Bennetts, 2014; Burton et al., 2010; McCaffery et al., 2018) has revealed large individual differences in ability in the neuro-typical population. These lie on a continuum ranging from developmental prosopagnosics (e.g., Knutson et al., 2011; McConachie, 1976), who display dysfunctional familiar face recognition; to super-recognisers, who display outstanding unfamiliar face recognition (e.g., Bate

et al., 2018; Bobak, Bennetts, et al., 2016; Bobak, Hancock, & Bate, 2016). These individual differences are heritable and face-specific, in that only weak correlations are found with non-face object recognition and other visual processing skills (e.g., Dunn et al., 2020; McCaffery et al., 2018; Royer et al., 2018; Verhallen et al., 2017; Wilhelm, Herzmann, et al., 2010; Wilmer, Germine, et al., 2010; Yovel et al., 2014).

Voice identification research has also revealed large individual differences in ability (Aglieri et al., 2017; Lavan, Burston, Ladwa, et al., 2019; Lavan, Merriman, Ladwa, et al., 2019; Mühl & Bestelmeyer, 2018). Indeed, phonagnosia, or the inability to recognise

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Applied Cognitive Psychology* published by John Wiley & Sons Ltd.

the voices of personally familiar people, has been described as an auditory analogue of prosopagnosia (Aglieri et al., 2017; Assal et al., 1976; Mühl et al., 2017; Van Lancker & Canter, 1982). As with faces, voice identification skills appear to be voice-specific, and are only moderately related to the recognition of non-voice auditory stimuli (i.e., bells, environmental sounds, musical instruments) (Aglieri et al., 2017; Hailstone et al., 2010; Von Kriegstein & Giraud, 2006). The current research is the first to explore the novel question of whether those who are exceptional at unfamiliar face recognition (super-recognisers) are also exceptional at voice recognition. A positive relationship would provide the basis for answering a second speculative question as to whether super-recognisers of voices might also exist.

This research has applied implications, as many policing and security operations require the identification of faces and voices. Trained forensic facial examiners outperform controls, and as a group, demonstrate equivalent levels of performance to untrained super-face-recognisers at simultaneous face matching (Phillips et al., 2018), suggesting some face identification tasks may be amenable to training (see Towler et al., 2017, for positive influence of training). However, it is not clear whether forensic face examination as a career may attract individuals with superior face recognition ability anyway. Furthermore, forensic facial examiner style training does not always produce performance improvements (e.g., Towler et al., 2014; Towler et al., 2019). Therefore, some police forces have deployed staff with naturally superior face processing abilities for roles drawing on their skills. Most successes from police super-face-recognisers are evident at the investigative stage, by, for instance, making multiple suspect identifications from CCTV for follow up by investigating teams (Davis et al., 2016, 2018, 2019; Robertson et al., 2016, for a review see Davis & Robertson, 2020). Individuals possessing exceptional unfamiliar voice recognition abilities may also be able to assist in investigatory contexts, for example, during phone surveillance, kidnapping, fraud, blackmail and counter-terrorism operations (e.g., *R v Kapikanya* [2015], Colman, 2016). Screening for exceptional unfamiliar voice recognition abilities may also become a useful component of recruitment to roles such as forensic voice examiners. Such individuals are sometimes required to provide expert voice comparison evidence in court. However, it would be very unlikely that such an expert would declare their superior voice recognition skills in court, as this would be irrelevant when providing opinion evidence, and may even be considered prejudicial to a case (see Davis et al., 2018 for a discussion on this point regarding super-face-recognisers).

## 1 | FAMILIAR AND UNFAMILIAR FACE AND VOICE RECOGNITION

Exploration of the relationship between individual differences in voice recognition and face recognition should also enhance theoretical understanding of the cognitive processes driving cross-modality human identification. A positive correlation might not necessarily be expected. Some 'pure' prosopagnosics and phonagnosics have been

documented who demonstrate specific face or voice deficits respectively, and whose existence suggests no overlap in face and voice capabilities (e.g., Assal et al., 1981; Liu et al., 2015, but see, for instance, Gainotti et al., 2003). Furthermore, faces and voices are mainly processed by dedicated face- and voice-selective brain areas (e.g., Assal et al., 1981; Kanwisher & Yovel, 2006; Pernet et al., 2015). However, activation in the fusiform face area, commonly associated with face processing, in response to a familiar speaker's voice alone (e.g., Blank et al., 2011; Von Kriegstein et al., 2005; von Kriegstein & Giraud, 2006), demonstrates cross-modal processing between voices and faces. As such, corresponding underlying cognitive mechanisms may drive voice and face recognition (e.g., Belin et al., 2011; see Yovel & Belin, 2013 for a review). Regardless, most people find identifying a voice harder than identifying a face (e.g., Brédart & Barsics, 2012; McAllister et al., 1993).

Functional models of familiar face (Bruce & Young, 1986) and familiar voice processing (Belin et al., 2004), have also proposed that person-specific stored face or voice representations respectively drive the identification of a known individual. With familiar faces, these viewpoint- and expression-independent representations enable reliable recognition even from poor-quality facial images (e.g., Bruce et al., 2001; Burton et al., 1999). Similarly, only a few idiosyncratic or salient voice cues are required for familiar voice recognition (Kreiman & Sittis, 2011). Indeed, familiar voice recognition is relatively robust to speech content generated variations in low-level acoustic features (e.g., Zäske et al., 2010), or vocal disguise (e.g., Hollien et al., 1982). Both face (Jeffrey et al., 2011; Valentine, 1991) and voice representations (Latinus & Belin, 2011; Lavan, Knight, & McGettigan, 2019; Papcun et al., 1989; Patel, 2008) also appear to be coded using corresponding modality-specific mechanisms (see also recent research consistent with this proposal: Lavan & Mcgettigan, 2019; Lavan et al., 2020). Nevertheless, although initial coding and recognition processes are physiologically independent, once face or voice representations are accessed (Belin et al., 2004), cross-modal familiar identity-related information and representations (i.e., name, face, voice) can be linked (for a recent review see Tsantani et al., 2019). Recent research results suggest that both modality-general and modality-specific mechanisms may both play a role in unfamiliar identity recognition. Johnson et al. (2020) revealed weak, but significant, correlations between voice and face sorting tasks, and, also, between identifying faces and voices as the "same identities" on a face and voice discrimination task, and at "telling people together" in face and voice sorting tasks. The authors suggest these results provide some evidence that common mechanisms underlie face and voice recognition, albeit additional modality-specific processes play a part.

Theories also propose that within each modality, unfamiliar and familiar identities are processed using qualitatively different mechanisms (for faces see Bruce & Young, 1986; Burton, 2013; Johnston & Edmonds, 2009; for voices see Kreiman & Sittis, 2011, 2013; Stevenage, 2018). Unfamiliar face recognition is governed by viewpoint- and expression-dependent pictorial codes. Small changes in appearance (i.e., facial expressions) or environmental conditions (e.g., lighting) can adversely impact recognition, even in tasks with no

memory demands (e.g., Davis & Valentine, 2009; Hill & Bruce, 1996; Jenkins et al., 2011). Corresponding mechanisms to those driving familiar and unfamiliar face processing have also been proposed for familiar and unfamiliar voice processing (Kreiman & Sidtis, 2011, 2013), with unfamiliar voice processing impacted by small variations in signal between learning and test (Lavan, Burston, & Garrido, 2019; Lavan, Merriman, Ladwa, et al., 2019). Basic acoustic patterns (e.g., pitch, vocal quality etc.) and features (e.g., masculinity, sharpness etc.) must be virtually identical across two clips for accurate unfamiliar voice recognition over even short retention intervals (e.g., Kreiman & Papcun, 1991). Despite these differences, there is usually a correlation between scores on familiar and unfamiliar face tests (Davis et al., 2016; Lander & Poyarekar, 2015; McCaffery et al., 2018, however, see Megreya & Burton, 2006), and sometimes between scores on familiar and unfamiliar voice tests (Van Lancker & Kreiman, 1987, although see Lavan, Burston, Ladwa, et al., 2019; Stevenage et al., 2020), indicative of common within-modality mechanisms, albeit effect sizes can be small.

## 2 | UNFAMILIAR FACE AND VOICE MEMORY AND MATCHING

Scores on different unfamiliar face identity tests normally moderately correlate ( $r \approx .5$ ) (e.g., McCaffery et al., 2018; Verhallen et al., 2017), suggesting that an underlying skill-based mechanism drives performance across the different tasks. However, case studies on super-recognisers have highlighted distinctions in performances between face memory and face matching. Indeed, some super-recognisers with exceptional unfamiliar face memory score relatively poorly on unfamiliar simultaneous face matching tests, and vice versa (e.g., Bennetts et al., 2017; Bobak, Bennetts, et al., 2016; Bobak, Hancock, & Bate, 2016; Davis et al., 2016). Unpublished online research conducted by the current authors (participants  $n \approx 42,000$ ) also supports a dissociation. While most who achieve super-recognition criteria (see Table 1 for definitions) on a face memory task also achieve super-recognition criteria at a face matching task (6.6% of total sample), 3.2% displayed superior face memory alongside typical-range face matching scores, while 1.8% achieved exceptional face matching alongside typical-range face memory scores. Similarly, some prosopagnosics with impaired face memory, display intact face matching skills (e.g., Bate & Bennetts, 2014; Dalrymple et al., 2014; Tardif et al., 2019). Corresponding dissociated performance patterns are found in phonagnosics, with some possessing impaired voice memory, but intact voice matching (Garrido et al., 2009; Xu et al., 2015). These studies imply that separate perception and memory-based mechanisms may drive task-based face processing and voice processing.

These results raise the intriguing possibility that if a common cross-modality mechanism drives voice and face identification skills, then some super-face-recognisers might also possess super-voice-recognition abilities. However, as previous studies within the face modality have shown that not all super-recognisers show superior

performance across memory and matching tasks, the current design allowed the exploration of whether there is consistency across face and voice performance in a specific task (i.e., if a face super-recogniser only excels at face matching and not face memory, they might be expected to excel at voice matching and not voice memory).

## 3 | AIMS AND HYPOTHESES

The aims of the current research were (1) to explore the relationship between face and voice recognition ability, (2) to examine whether super-recognition status generalises across face and voice modalities, and (3) to assess whether dissociated performance patterns between face memory and matching tests would also be found with voice memory and matching tests. Accordingly, participants completed an unfamiliar voice matching test (Bangor Voice Matching Test [BVMT]; Mühl et al., 2017), a new familiar voice memory test (Famous Voice Recognition Test [FVRT]), and an unfamiliar voice memory test (Glasgow Voice Memory Test [GVMT]; Aglieri et al., 2017), which contains a bell recognition component, allowing the isolation of voice-specific memory scores. All participants had previously completed an unfamiliar face memory test (Cambridge face memory test: Extended, CFMT+, Russell et al., 2009), commonly used to allocate super-recognisers to research groups, and an unfamiliar face matching test (Glasgow face matching test [GFMT], Burton et al., 2010). Four groups were formed: Superior-Face-Identifiers (SF-Identifiers), who had registered outstanding scores on the CFMT+, and GFMT, indicative of superior face memory and face matching, Superior-Face-Recognisers (SF-Recognisers); exceptional at face memory only, Superior-Face-Matchers (SF-Matchers); exceptional at face matching only; and Controls with typical-range scores on both face tests.

Assuming that face and voice recognition are underpinned by a common cross-modality mechanism (e.g., Belin et al., 2004; Tsantani et al., 2019), it was hypothesised that a positive correlation would be found between voice and face recognition test performances. More specifically, it was predicted that participants with exceptional face matching (SF-Identifiers, SF-Matchers) would outperform those with typical-range face matching (SF-Recognisers, Controls) on the BVMT (Mühl et al., 2017). Likewise, participants with exceptional face memory (SF-Identifiers, SF-Recognisers) would outperform those with typical-range face memory (SF-Matchers, Controls) on the GVMT (Aglieri et al., 2017) and the FVRT. Finally, and consistent with research finding voice-specific effects, no between-group differences were expected on the bells stimuli of the GVMT, supporting propositions that voice-specific processing ability is independent of processing of other auditory stimuli (e.g., Aglieri et al., 2017).

## 4 | METHOD

This study received ethical approval from the University of Greenwich Research Ethics Committee (17.4.5.16), and the Ethics Committee of the University of Southampton School of Psychology.

## 5 | DESIGN

The study primarily employed a between-groups design (SF-Identifiers, SF-Recognisers, SF-Matchers, Controls), based on previous CFMT+ (Russell et al., 2009), and GFMT (Burton et al., 2010) scores. Performances of groups were compared on the BVMT (Mühl et al., 2017), the GVMT (Aglieri et al., 2017), and a bespoke FVRT. Outcomes were assessed by hit rates (HRs) (correct identifications of voices as old (GVMT), same decisions (BVMT), or familiar decisions (FVRT)); correct rejection rates (CRs) (correct identifications of voices as being new, different, unfamiliar respectively), and measures of sensitivity ( $d'$ ), and response bias ( $C$ ) (Green & Swets, 1966; Macmillan & Creelman, 1991, see data analyses section). A correlational component also examined relationships between-test performances.

### 5.1 | Participants

From an existing volunteer University of Greenwich database of approximately 42,000 participants, 13,263 participants were invited to participate as they met group inclusion criteria (see Table 1). All had previously provided consent to be contacted to take part in future research studies, and for access to previously recorded test scores (CFMT+, GFMT). Six hundred and five finished all tests.

Participants were excluded if they failed to input the correct participant ID code ( $n = 9$ ), if they failed to recognise any of the celebrity voices on the FVRT ( $n = 6$ ), or if they scored below chance levels on the BVMT and GVMT ( $n = 61$ ). The latter was mostly due to participants using a problematic browser despite recommendations to avoid. This issue was experienced across groups and, although likely to reduce effect sizes, conclusions would unlikely be different. The final sample consisted of 529 participants with a mean age of 36.9 years ( $SD = 11.8$ , Range = 18–76, 64% female).

Table 1 presents group inclusion criteria. The SF-Identifier group possessed outstanding face memory (CFMT+ threshold:  $\geq 93$  out of 102, equivalent to that estimated to be achieved by the top 2% of the population: Belanova et al., 2018), and outstanding unfamiliar face matching ability (GFMT threshold: maximum score of 40 out of 40). The Control group consisted of individuals displaying typical-range performances (i.e., within 1  $SD$  of previous norms) on the CFMT+ (58–83/102: Bobak, Pampoulov, & Bate, 2016) and GFMT (28–36/40: Burton et al., 2010; Robertson et al., 2016). The SF-Recogniser and

SF-Matcher groups respectively displayed superior performance on the CFMT+ only, or the GFMT only, while producing typical-range scores on the alternative test.

## 6 | MATERIALS

### 6.1 | Bangor Voice Matching Test (BVMT)

The BVMT is an unfamiliar voice discrimination task consisting of 80 speaker pairs (40 male) (Mühl et al., 2017). Instructions are displayed on-screen and the test is self-paced. For each trial, a central fixation cross displays for 800 ms before stimulus onset. Participants then view two speaker icons presented side-by-side on a monitor screen. Clicking on each icon generates a different audio track in which the two speakers voice one different syllable each. Participants are able to click between the icons; thus, presentation is dichotomous. Verbalisations are constructed from consonant-vowel-consonants (CVC; e.g., had, hed etc.) and vowel-consonant-vowels (VCV; e.g., aba, ibi etc.), and configured with a minimum of eight syllable pairs (CVC-CVC, VCV-VCV, VCV-CVC, or CVC-VCV). Participants can listen to each audio track multiple times, before either selecting the Same Person or Two Different People response option. The 'next trial' icon then appears. The test takes approximately 10–15 minutes to complete. The test yields data in the form of HRs correct decisions that voice pairs belong to the same speaker (40 trials); and correct rejections (CRs) that voice pairs belong to different speakers (40 trials), allowing calculation of sensitivity ( $d'$ ) and response bias or criterion ( $C$ ).

### 6.2 | Glasgow Voice Memory Test (GVMT)

The GVMT is a four-stage test examining unfamiliar voice memory for 16 voices (8 male, 8 female) and memory for 16 bells (Aglieri et al., 2017). In stage 1 (encoding voices), eight voices (4 male) utter the French vowel/a/three times each. Almost immediately in stage 2 (recognising voices), the same eight voices are randomly presented together with eight new voices (4 male). All utter the same vowel sound/a/as before. Participants respond old voice or new voice to indicate whether each voice was previously heard. The same old/new design is employed for stages 3 and 4 (bells) which consists of eight old and eight new bells. The authors note that the sound clips for bells

**TABLE 1** Inclusion criteria for participant groups

	<i>n</i>	CFMT+ Score (out of 102)	GFMT Score (out of 40)	<i>M</i> age ( <i>SD</i> ) Years	Age range Years	Proportion female %
SF-Identifiers	165	$\geq 93$	40	35.7 (9.4)	19–68	72.1
SF-Recognisers	89	$\geq 93$	28–36	38.4 (10.7)	19–73	55.1
SF-Matchers	41	58–83	40	34.2 (13.2)	19–70	70.7
Controls	234	58–83	28–36	37.7 (13.3)	18–76	61.1

Note: From the database of 42,000 volunteer participants all SF-Identifiers ( $n = 2778$ ), SF-Recognisers ( $n = 1368$ ), SF-Matchers ( $n = 768$ ) and Controls ( $n = 8349$ ) were invited, although far fewer were eligible (at least 50% reside outside the UK).

( $M \approx 1190$  ms) are longer in duration than the voice clips ( $M \approx 487$  ms), while there is more variation in the pitch and timbre of the bells, both of which might aid recognition. The test takes approximately 10–15 minutes to complete. HRs (correct 'old' decisions), CRs (correct 'new' decisions), sensitivity ( $d'$ ) and criterion (C) were again calculated.

### 6.3 | Famous Voice Recognition Test (FVRT)

This 38-trial bespoke test contains voice samples (30 s) of 28 famous and 10 unfamiliar individuals, providing information about topics unlikely to assist in identification (i.e., charities).<sup>1</sup> The test starts with a practice trial to guide participants on the procedure for this test. For this, and each subsequent trial, participants listen to a target voice clip and respond 'yes' or 'no' as to whether they recognise the speaker. If they respond 'yes,' they type brief individuating information about the speaker to support their decision. If they respond 'no,' they type 'UN' (for unknown), before selecting 'next.' A photo of the target is then displayed together with some individuating information about the target, and participants respond 'yes' or 'no,' as to whether they recognise the target. Participants also rate their familiarity with that target on a 0 (not at all)–7 (highly familiar) Likert-scale. Participants then move onto the next trial.

The test takes approximately 20 minutes to complete. Participant responses were checked for correct identifications of famous voices (providing a correct name or clear, unambiguous description), once by the researcher and again by an independent second rater. A check of inter-rater reliability yielded a mean concordance rate of 99.5% for all 28 famous trials.

Conditionalised Naming HRs (CNHR) were calculated by computing the number of correct identifications of famous voices (providing a correct name or clear, unambiguous description), but only if the participant also responded that the target photo was familiar. A rating of '≥1' on the 0–7 Likert-scale above formed a minimum criterion for determining familiarity. If participants were unfamiliar with a target's photo (e.g., response of zero), the response to that target's voice was excluded from analyses. For example, if 10 target photos were familiar to a participant, and they provided correct individuating information to eight of those voices, this would yield a CNHR of .80. CRs were correct UN (unknown) responses. Sensitivity ( $d'$ ) and bias (C) were also calculated from CNHRs and CRs.

Pilot tests: Online super-face-recogniser research using similar public recruitment strategies as the current study (Belanova et al., 2018; Satchell et al., 2019) has tended to attract higher-ability participants than laboratory studies (Bate et al., 2018; Bobak, et al., 2016). This is a likely consequence of online participants not responding to adverts if they know their abilities are poor. To examine whether a similar recruitment bias would be found in the current research, psychology students ( $N = 75$ ) from two UK universities were tested and received course credit for their time. Similar exclusion criteria were applied to students as those reported in the participants section ( $n = 13$ ), leaving a total student sample of 62 (59 female, aged

18–39 years,  $M = 20.19$ ,  $SD = 4.00$ ). These participants were expected to be representative of typical-range ability and were tested online. This allowed for a direct comparison of performances with the control group described above.

Although no effects reached significance on the FVRT ( $p > .05$ ), analysis of most other measures revealed better performance in controls compared to students. Indeed, a series of independent-measures  $t$ -tests on the main outcome,  $d'$ , revealed that controls significantly outperformed students on the BVMT ( $p = .032$ ), and the bells section of the GVMT ( $p < .001$ ). No significant differences were found for  $d'$  scores on the voice section of the GVMT ( $p > .05$ ) between students and controls. This suggests that even though controls scoring in the typical-range on the CFMT+ and GFMT were invited from the database (i.e., with an upper limit on test performances—unlike the students), better than typical performers at voice processing tended to take part.

## 7 | PROCEDURE

All voice tests were administered via Qualtrics. Participants were invited via email and were provided with a 9-digit anonymised code linked to previous CFMT+ and GFMT data. They were asked to avoid using a specific named browser, as audio play was sometimes problematic. Many failed to heed the warning, hence the exclusions reported above.

After providing informed consent, and permission to access previous face test scores, participants entered their anonymous code. They were asked to wear headphones and completed the BVMT, GVMT, and the FVRT in that order. Total test time was 40–50 minutes. Breaks were encouraged between tests. All participants were debriefed.

Data Analyses: It is noteworthy that differences exist between the psychological mechanisms that drive HRs (recognition of previously encountered stimuli) and decision-making processes associated with CRs (recognising that stimuli have not been encountered previously) (Bate et al., 2018; Belin et al., 2004, 2011; Bruce & Young, 1986; Kreiman & Sidtis, 2011; Megreya & Burton, 2006). It was therefore important to measure HRs and CRs, as well as the independent signal detection theory measures of sensitivity ( $d'$ ) and response bias (C) (Green & Swets, 1966; Macmillan & Creelman, 1991). Sensitivity ( $d'$ ) measures reflect levels of discrimination between old and new (in memory tests), same and different (in matching tests), and familiar and unfamiliar (in familiarity tests). A high, positive value of  $d'$  indicates good discrimination. Response bias or criterion (C), is independent of sensitivity, and measures the tendency to respond one way or another under conditions of uncertainty. A positive value of C indicates a conservative bias or a tendency to respond 'new' in memory tests—'different' in matching tests—'unfamiliar' in familiar tests. Negative values of C indicate a liberal bias or a tendency to respond 'old,' 'same,' or 'familiar,' respectively. To calculate  $d'$  and C, HRs, and false alarm rates of 1.0 and 0.0 were adjusted using the formulas  $1-1/(2 N)$ , and  $1/(2 N)$  respectively to prevent summation to infinity.



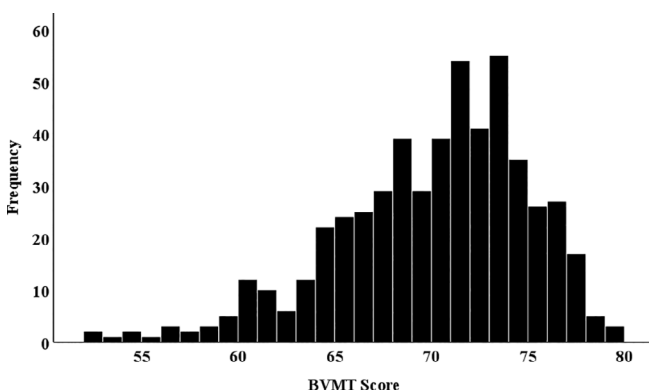
For some SDT-based research, reporting  $d'$  and  $C$  only, and not all four measures, has been sufficient to reveal important effects. However, as with some person-recognition research (e.g., Bate et al., 2018; Belanova et al., 2018), here such a strategy would have obscured important effects, particularly those involving CRs. Therefore, all four statistics are reported.

To compare the outcomes of the four groups (SF-Identifiers, SF-Recognisers, SF-Matchers, Controls), a series of one-way between-group ANOVAs were conducted on HRs, CRs, sensitivity ( $d'$ ), and bias ( $C$ ) on the BVMT and FVRT. Four 4 (group)  $\times$  2 (stimulus-type) mixed ANOVAs were conducted on the GVMT outcomes to account for the separate voices and bells components, and to assess voice-specific group differences. Under applications of non-normally distributed data, ANOVAs are relatively robust (Schmider et al., 2010). Alpha levels were set at .05. Due to uneven group sizes, for post-hoc analyses, the Tukey-Kramer test was employed (Hayter, 1984).

To test the modality-general (voices and faces) hypotheses that task-specific (memory vs. matching) superior face ability would transfer to the task-specific (memory vs. matching) voice tests, groups were collapsed into exceptional and typical-ability groups. This increased statistical power to identify effects. For the voice matching test, SF-Identifiers and SF-Matchers formed an Exceptional-Matcher group, while a Typical-Matcher group was formed from SF-Recognisers and Controls. For the voice memory tasks, SF-Identifiers and SF-Recognisers formed an Exceptional-Recogniser group, while SF-Matchers and Controls formed a Typical-Recogniser group. Independent-measures  $t$ -tests compared performances between super and typical-ability groups on all outcomes.

## 8 | RESULTS

**Bangor Voice Matching Test:** Mean HRs varied from .30–1.00 ( $M = .91$ ,  $SD = .08$ ) while mean CRs ranged from .33–1.00 ( $M = .83$ ,  $SD = .11$ ). No participant achieved a score of 100% on the BVMT. Figure 1 illustrates the range of scores and Figure 2 summarises mean group performances on each measure (HRs, CRs,  $d'$ ,  $C$ ).



**FIGURE 1** Range of scores for the Bangor Voice Matching Test (BVMT) (Shapiro-Wilk (529) = .960,  $p < .001$ ; Skewness =  $-.751$ , kurtosis = .493)

A series of one-way ANOVAs revealed significant group main effects for HRs ( $F(3, 525) = 6.78$ ,  $p < .001$ ,  $\eta_p^2 = .037$ ), CRs ( $F(3, 525) = 3.76$ ,  $p = .011$ ,  $\eta_p^2 = .021$ ), and  $d'$  ( $F(3, 525) = 18.26$ ,  $p < .001$ ,  $\eta_p^2 = .094$ ), but not for  $C$  ( $F(3, 525) < 1$ ). As expected, SF-Identifiers significantly outperformed controls at HRs ( $p < .001$ ), CRs ( $p = .010$ ), and  $d'$  ( $p < .001$ ); and SF-Recognisers at  $d'$  ( $p = .002$ ). SF-Matchers also significantly outperformed controls at  $d'$  ( $p = .003$ ) and marginally at HRs ( $p = .063$ ).

Table 2 summarises the mean performances of the combined Exceptional-Matcher (SF-Identifiers, SF-Matchers) and Typical-Matcher groups (SF-Recognisers, Controls), as well as the results of  $t$ -tests addressing the hypothesis that those with superior face matching skills would also possess superior voice matching abilities. Meeting predictions, the Exceptional-Matcher group significantly outperformed the Typical-Matcher group on HRs, CRs, and  $d'$  on the BVMT. There were no criterion differences.

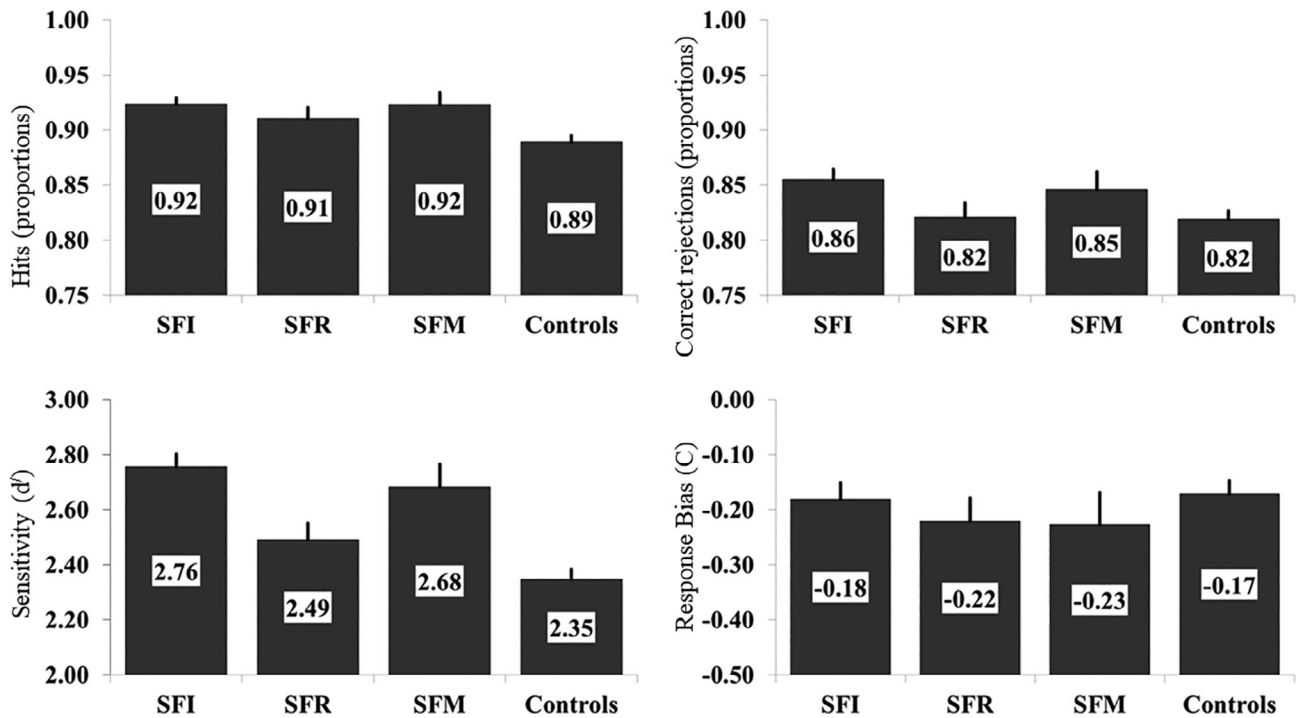
**Glasgow Voice Memory Test:** For voices, mean HRs (correct 'old' responses) varied from .25–1.00 ( $M = .78$ ,  $SD = .16$ ), while mean CRs (correct 'new' responses) ranged from .13–1.00 ( $M = .68$ ,  $SD = .16$ ). For bells, overall mean HRs varied from .38–1.00 ( $M = .83$ ,  $SD = .14$ ), while mean CRs ranged from .13–1.00 ( $M = .70$ ,  $SD = .15$ ). Overall, four participants achieved scores of 100% on the voice section of the test; another five achieved 100% on the bells section of the test. Figure 3a,b illustrate the range of scores on the voice and bell section of the GVMT, respectively.

Figure 4 summarises the mean performances for voices and bells for each group. To address the hypothesis that effects would be voice-specific, the results of a 4 (group: SF-Identifier, SF-Recogniser, SF-Matcher, Control)  $\times$  2 (stimulus-type: voices, bells) mixed ANOVA revealed significant stimulus-type main effects for HRs,  $F(1, 525) = 16.22$ ,  $p < .001$ ,  $\eta_p^2 = .030$ , and  $d'$ ,  $F(1, 525) = 15.49$ ,  $p < .001$ ,  $\eta_p^2 = .029$ . Bells were better recognised and discriminated than voices, consistent with the expected bell-superiority pattern. The group main effect was not significant for any outcome ( $p > .2$ ).

A significant interaction was revealed for CRs only,  $F(3, 525) = 2.94$ ,  $p = .033$ ,  $\eta_p^2 = .017$ . Two one-way between-group ANOVAs on CRs revealed significant effects for voices,  $F(3, 525) = 4.02$ ,  $p = .008$ ,  $\eta_p^2 = .022$ , but not bells,  $F(3, 525) < 1$ , supporting the hypothesis that between-group effects would be voice-specific only.

Subsequent Tukey-Kramer post-hoc tests on the CRs for voices revealed that, as anticipated, SF-Identifiers outperformed controls ( $p = .011$ ) at identifying that a voice had not been heard before. Surprisingly, SF-Identifiers ( $M = .71$ ) also outperformed SF-Recognisers ( $M = .65$ ) ( $p = .028$ ), who surprisingly did not outperform controls ( $M = .66$ ). No other between-group differences were significant ( $p > .2$ ).

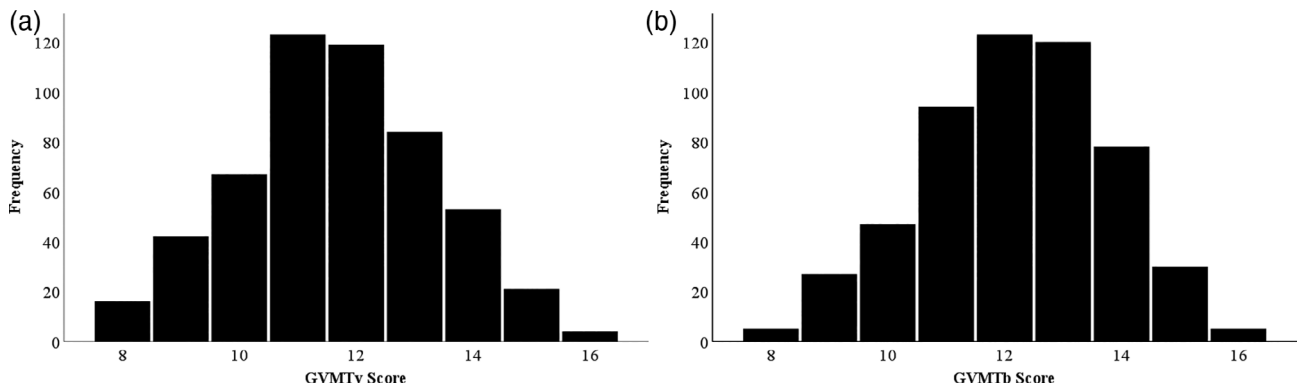
To address the modality-general hypothesis that those with superior face memory skills would also possess superior voice memory,  $t$ -tests found that the Exceptional-Recognisers (SF-Identifier, SF-Recogniser) significantly outperformed the Typical-Recognisers (SF-Matchers, Controls) at  $d'$ , and CRs (marginally) (see Table 2). There were no significant effects with bells, further supporting the voice-specific hypothesis.



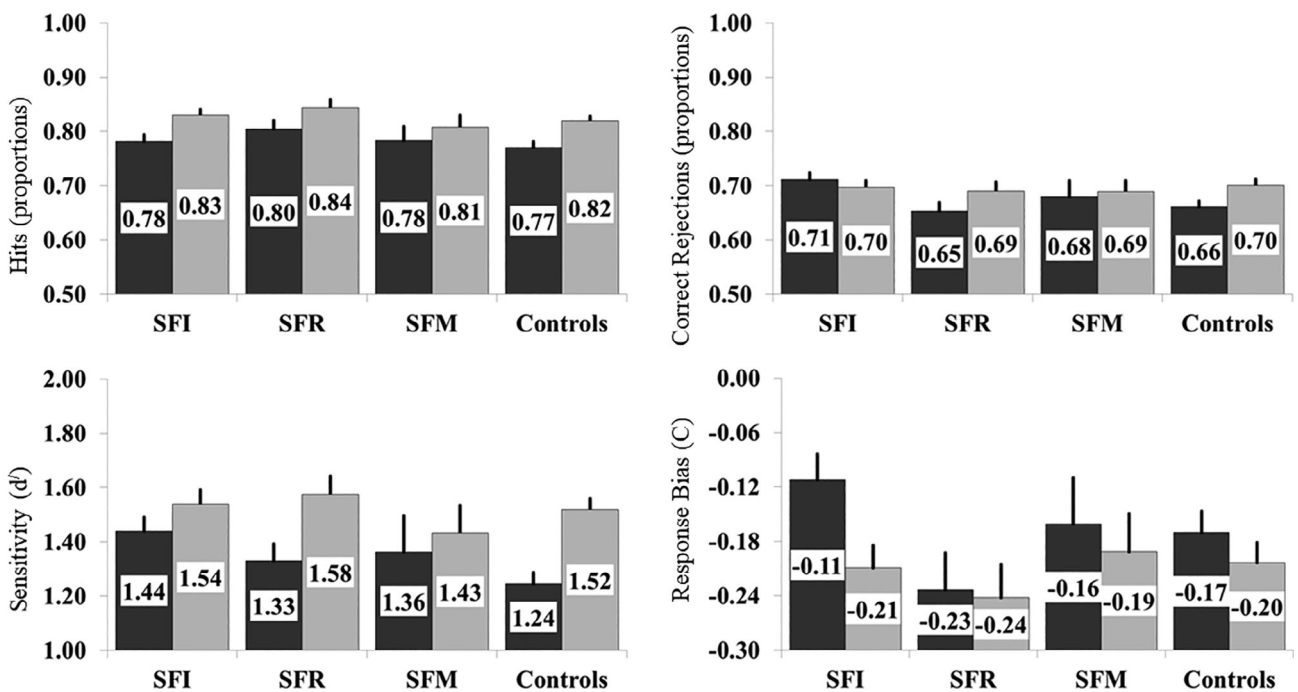
**FIGURE 2** From top left clockwise, Bangor Voice Matching Test HRs, CRs, C, and  $d'$  for SF-Identifiers (SFI), SF-Recognisers (SFR), SF-Matchers (SFM) and controls. Error bars denote standard errors of the mean (+1 SEM)

**TABLE 2** Results for t-tests comparing the collapsed super groups with typical-ability groups (see text for explanation)

	Combined super-groups		Combined typical-ability groups		df	t	d	p
	M	SD	M	SD				
<b>Bangor Voice Matching Test</b>								
	Exceptional-matchers		Typical-matchers					
HRs	0.92	0.07	0.90	0.09	527	3.98	0.37	<.001
CRs	0.85	0.11	0.82	0.11	527	3.34	0.30	.001
$d'$	2.74	0.58	2.39	0.56	527	7.05	0.63	<.001
C	-0.19	0.39	-0.19	0.38	527	-0.14	0.01	>.2
	Exceptional-recognisers		Typical-recognisers					
<b>Glasgow Voice Memory Test</b>								
Voices								
HRs	0.79	0.15	0.77	0.16	527	1.26	0.11	>.2
CRs	0.69	0.16	0.66	0.16	527	1.93	0.17	.054
$d'$	1.40	0.62	1.26	0.67	527	2.47	0.21	.014
C	-0.15	0.37	-0.17	0.35	527	0.45	0.04	>.2
Bells								
HRs	0.84	0.14	0.82	0.14	527	1.48	0.13	.138
CRs	0.69	0.16	0.70	0.15	527	-0.39	0.03	>.2
$d'$	1.55	0.64	1.51	0.62	527	0.85	0.07	>.2
C	-0.22	0.33	-0.19	0.33	527	-1.18	0.10	>.2
<b>Famous Voice Recognition Test</b>								
HRs	0.31	0.23	0.24	0.22	527	3.58	0.31	<.001
CRs	0.90	0.14	0.91	0.13	527	-0.59	0.05	>.2
$d'$	0.80	0.91	0.58	0.93	527	2.73	0.24	.007
C	1.01	0.54	1.15	0.53	527	-3.08	0.27	.002



**FIGURE 3** (a) Range of scores for the voice section of the Glasgow Voice Memory Test (GVMT) (Shapiro–Wilk (529) = .966,  $p < .001$ ; Skewness = .024, kurtosis =  $-.362$ ). (b) Range of scores for the bells section of the Glasgow Voice Memory Test (GVMT) (Shapiro–Wilk (529) = .962,  $p < .001$ ; Skewness =  $-.183$ , kurtosis =  $-.337$ )



**FIGURE 4** From top left clockwise, Glasgow Voice Memory Test HRs, CRs, C, and  $d'$  for SF-Identifiers (SFI), SF-Recognisers (SFR), SF-Matchers (SFM) and controls. Dark bars represent voice section and grey bars represent bell section. Error bars denote standard errors of the mean (+1 SEM)

*Famous Voice Recognition Test:* Unadjusted famous voice recognition HRs (not adjusted for familiarity) varied from .0–.93 ( $M = .21$ ,  $SD = .19$ )—see Figure 5 for the range of scores not adjusted for familiarity. CNHRs ranged from 0–1.00 ( $M = .28$ ,  $SD = .23$ ). CRs however, were close-to-ceiling and ranged from 0–1.00 ( $M = .90$ ,  $SD = .13$ ). Figure 6 summarises mean group performances (CNHRs, CRs,  $d'$ , C).

There were significant group main effects for CNHRs ( $F(3, 525) = 5.73$ ,  $p < .001$ ,  $\eta_p^2 = .032$ ),  $d'$  ( $F(3, 525) = 2.94$ ,  $p = .033$ ,  $\eta_p^2 = .017$ ) and C ( $F(3, 525) = 4.94$ ,  $p = .002$ ,  $\eta_p^2 = .027$ ), but not for CRs ( $F(3, 525) < 1$ ). SF-Identifiers significantly outperformed controls at CNHRs, and the discrimination ( $d'$ ) of famous and non-famous voices. These effects were partly a consequence of a significantly more conservative response bias by controls, in that they were more

cautious than SF-Identifiers in attributing identity, with a greater tendency to report 'unknown' thus increasing CRs.

Table 2 provides further support for the modality-general hypothesis in that the Exceptional-Recognisers (SF-Identifiers, SF-Recognisers) significantly outperformed the Typical-Recognisers (SF-Matchers, Controls) in terms of CNHRs, and  $d'$  on the FVRT. The Typical-Recognisers group also displayed a significantly more conservative response bias than the Exceptional-Recognisers ( $p = .002$ ), similar to that described above.

*Individual analyses:* To further examine the relationship between voice and face processing abilities, correlational analyses were conducted on all test outcomes. The Benjamini-Hochberg (Benjamini & Hochberg, 1995) adjustment, set at .10, was applied to protect against



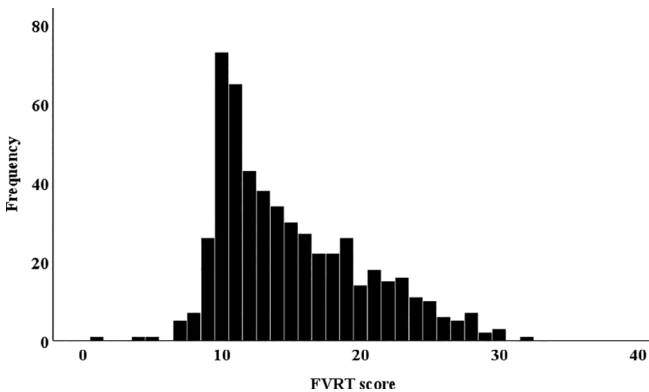
Type-1 errors associated with multiple comparisons. Results of the correlations on HRs, CRs, and C for the three voice tests, and CFMT+ and GFMT accuracy can be found in the supplementary data (S1). Table 3 reports the sensitivity ( $d'$ ) correlations only. This measure also generated the strongest between-group effects described above. Patterns for hits, CRs and C in the supplementary data (S1) mostly matched those reported for  $d'$ , although not all comparisons reached significance.

As predicted, the results replicated the well-established strong correlation between scores on the CFMT+ and the GFMT. CFMT+ scores were also significantly correlated, albeit less strongly (Cohen, 1988), with  $d'$  on all voice tests, but not with bells on the

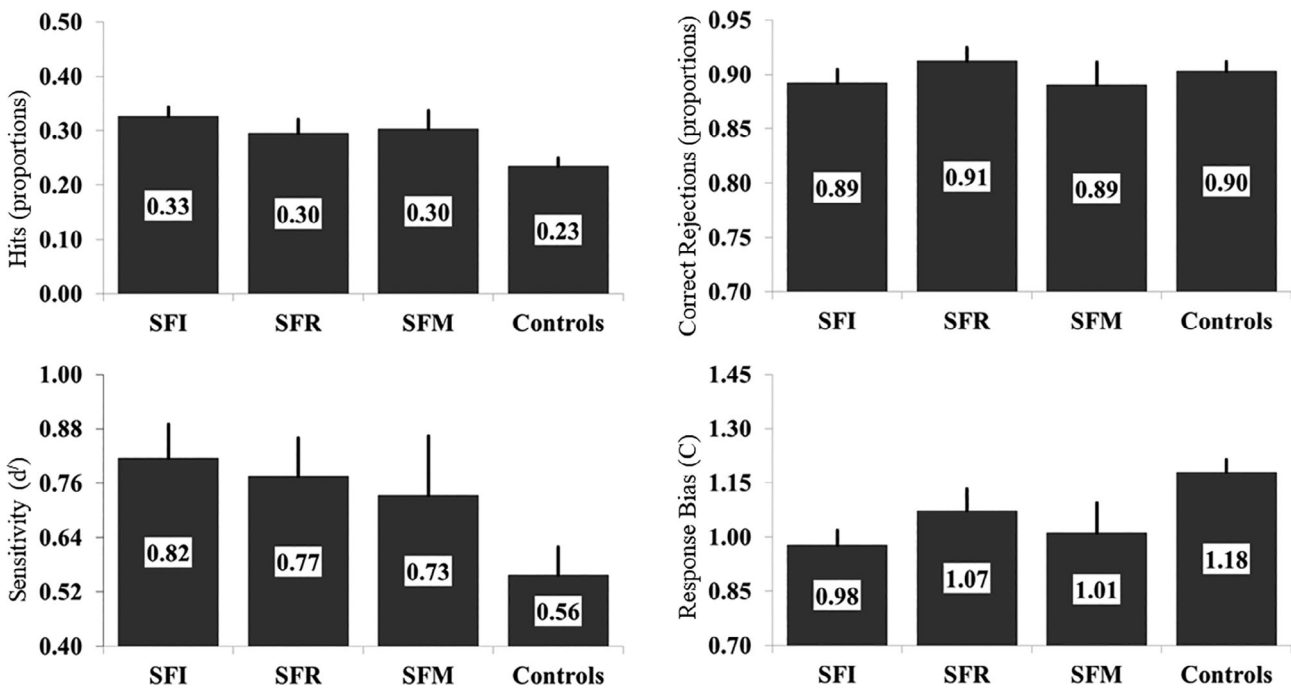
GVMT. This suggests underlying face and voice memory mechanisms may be related, but that they are not related to memory for bells. However, in support of previous literature (Aglieri et al., 2017), significant positive correlations were found between voice and bell sections of the GVMT, which may be due to the identical task demands on the two components of this test, allowing for test-specific strategies to be employed. Similarly, as anticipated, scores on the GFMT were significantly moderately correlated with the BVMT ( $d'$ ), consistent with the between-group test results on this measure. The GFMT also significantly, but weakly, correlated with voice recognition on the GVMT and FVRT, but not with bell recognition—once again supporting our voice-specific hypothesis. As with correlations between scores on the CFMT+ and GFMT, correlations between the GVMT and BVMT were significant, albeit contrastingly weak, replicating similar effects found previously (Mühl et al., 2017). Likewise, significant correlations between the FVRT and the other voice tests were weak.

*Potential Super-Voice-Recognisers:* Table 4 provides a tally of potential 'super-voice-recognisers'. With face research, a criterion of 2 SD above control means has normally been applied to identify super-recognisers in the approximate top 2% of the population. Here, a liberal criterion of 1.5 SD for each test was employed allowing borderline cases to be included (see also Bate et al., 2019). This was also necessary as voice test scores, particularly on the BVMT, were often close-to-ceiling (see also Mühl et al., 2017), while numbers of stimuli were relatively low on all tests.

Although there were between-test inconsistencies, as would be predicted given the results reported above, the SF-Identifier group were most likely to be classified as potential super-voice-recognisers on each test, while controls were the least likely. Some SF-Identifiers



**FIGURE 5** Range of scores for the Famous Voice Recognition Test (FVRT) (unadjusted for familiarity) (Shapiro–Wilk (529) = .930,  $p < .001$ ; Skewness = .772, kurtosis =  $-.068$ )



**FIGURE 6** From top left clockwise Famous Voice Recognition Test (FVRT) CNHRs, CRs, C, and  $d'$  for SF-Identifiers (SFI), SF-Recognisers (SFR), SF-Matchers (SFM) and controls. Error bars denote standard errors of the mean (+1 SEM)

**TABLE 3** Pearson's correlation coefficients between primary test outcomes

	GFMT score	BVMT $d'$	GVMT:V $d'$	GVMT:B $d'$	FVRT $d'$
Cambridge Face Memory Test (CFMT+)	.59**	.25**	.12**	.05	.15**
Glasgow Face Matching Test (GFMT)	—	.30**	.11*	-.03	.11*
Bangor Voice Matching Test (BVMT)	—	—	.20**	.17**	.08*
Glasgow Voice Memory Test: voices (GVMT:V)	—	—	—	.19**	.10*
Glasgow Voice Memory Test: Bells (GVMT:B)	—	—	—	—	.09*

Note: FVRT = Famous Voice Recognition Test.

\*Benjamini-Hochberg corrected ( $p < .05$ ) for multiple comparisons (false discovery rate = 0.10).

\*\*Benjamini-Hochberg corrected ( $p < .001$ ) (false discovery rate = 0.10).

**TABLE 4** Potential super-voice-recognisers based on a liberal criteria of 1.5 SDs above control means for the Bangor Voice Matching Test (BVMT), the Glasgow Voice Memory Test (GVMT) voices (V) and bells (B); and the Famous Voice Recognition Test. Percentages represent proportion within each group achieving criteria

	BVMT $d'$	GVMT:V $d'$	GVMT:B $d'$	FVRT $d'$
Total ( $n = 529$ )	71 (13.4%)	78 (14.7%)	35 (6.6%)	71 (13.4%)
Groups				
SF-Identifier ( $n = 165$ )	41 (24.8%)	28 (17.0%)	13 (7.9%)	28 (17.0%)
SF-Recogniser ( $n = 89$ )	9 (10.1%)	10 (11.2%)	8 (9.0%)	13 (14.6%)
SF-Matcher ( $n = 41$ )	6 (14.6%)	12 (29.2%)	1 (2.4%)	6 (14.6%)
Controls ( $n = 234$ )	15 (6.4%)	28 (12.0%)	13 (5.6%)	24 (10.3%)

exceeded the 1.5 SD criterion on two ( $n = 9$ ) and three voice tests ( $n = 2$ ). A few SF-Recognisers ( $n = 5$ ), SF-Matchers ( $n = 2$ ), and controls ( $n = 6$ ) also exceeded criteria on two voice tests; while one SF-Matcher exceeded criteria on all three voice tests, all indicative of possible super-voice-recognition ability.

The pattern for the bells component of the GVMT provides a strong contrast to the three tests of voice identification. As would be expected given the voice-specific predictions and results reported above, far fewer participants achieved the criteria for 'super-recognition' of bells. These results support the main hypothesis that superior face identification skills would primarily transfer to the identification of voices only.

## 9 | DISCUSSION

The current study aimed to explore the relationship between face and voice-specific recognition ability, in order to examine whether super-recognition status generalises across face and voice modalities, and to determine whether dissociated performance patterns between face memory and matching tests would be found in the same participants when completing voice memory and matching tests. As predicted, a positive relationship was found between performances on tests of face and voice identity processing. Participants with pre-assessed superior unfamiliar face matching (GFMT, Burton et al., 2010) and/or face memory ability (Cambridge Face Memory Test: Extended; CFMT+, Russell et al., 2009) tended to possess superior unfamiliar voice matching (BVMT, Mühl et al., 2017), and/or unfamiliar (GVMT, Aglieri et al., 2017), and familiar voice memory ability (FVRT) as well. These

findings are consistent with recent literature suggesting modality-general integration of voices and faces (Tsantani et al., 2019). Correlations between voice and face tests were mainly small to moderate, a potential consequence of weak voice test discriminatory power. Nevertheless, these results provide some support to models of voice and face recognition suggesting that voice and face identity processing may be driven by common underlying mechanisms (Belin et al., 2004, 2011; Bruce & Young, 1986). Indeed, effects were primarily voice-specific, as superior face identification ability did not transfer to superiority at bells recognition on the GVMT. Indeed, bell stimuli were roughly equally recognised by each face-ability-based group. These effects are unlikely to be due to ceiling effects, as HRs and CRs for bells were both below .85, albeit significant main effects on hits and sensitivity demonstrates that the bells stimuli are easier to recognise and to discriminate between than voices. These results are consistent to previous research using the same stimuli, and Aglieri et al. (2017) explains this by noting that the bell stimuli are played for longer than the vocal stimuli, and possess greater acoustic variability (e.g., pitch and timbre), whereas the voices in the test tend to have highly similar acoustic properties as they all utter a single vowel sound. Nevertheless, significant correlations between the three voice tests, albeit weaker than those between the face tests, also provide tentative support that a common within-modality mechanism may underlie the processing of familiar and unfamiliar voices, a proposition that requires further research.

Perhaps not surprisingly, being originally designed to measure deficits in ability, close-to-ceiling effects were found on the BVMT. Performances were contrastingly worse on the FVRT, in which scores were conditionalised on being able to recognise a photograph of each

celebrity. Overall, participants correctly recognised fewer than one-third of the voices belonging to known faces. These results support previous research demonstrating the relative difficulty of voice recognition compared to face recognition (Brédart & Barsics, 2012; McAllister et al., 1993, but see, for instance, McCaffery et al., 2018 for similar scores on a famous face test). Nevertheless, CR rates on the FVRT were close-to-ceiling for all groups, driven by a bias to respond 'unknown' to voices, regardless of whether that celebrity was familiar or not.

Despite the issues associated with the test designs, between-group analyses also provide support for the proposal that superior task face processing ability transfers to the voice modality. The combined Exceptional-Matcher group (SF-Identifiers, SF-Matchers) outperformed the Typical-Matcher group (SF-Recognisers, Controls) on the BVMT. SF-Identifiers, possessing exceptional face memory and face matching skills, also made significantly more correct matched and mismatched decisions than controls. Compared to controls, SF-Identifiers additionally made significantly more hits of famous voices, and more CRs of unfamiliar voices on the GVMT. Likewise, the combined Exceptional-Recogniser group (SF-Identifiers, SF-Recognisers) made significantly more hits of famous voices than Typical-Recognisers (SF-Matchers, Controls). They also displayed significantly better discrimination of unfamiliar voices on the GVMT.

It should be noted that SF-Identifiers' superior GVMT performances were driven by CRs of previously unheard voices only. This might suggest SF-Identifiers' face superiority does not transfer to the recognition of previously heard voices but to the rejection of new voices. However, recent research suggests that discriminating between two different unfamiliar voices or "telling (different) voices apart," is more accurate than performance when generalising across within-speaker variability, or "telling (same) voices together" (e.g., Lavan, Burston, & Garrido, 2019; Lavan, Merriman, Ladwa, et al., 2019; Stevenage et al., 2020; see Jenkins et al., 2011 for similar effects with faces). As such, an enhanced ability to discriminate between two different voices appears to be the driver of the SF-Identifier effects found here. It is also possible that the GVMT does not generate enough variability in HRs to distinguish between participants with exceptionally good and moderate ability levels. Ability-based differences in hits might be revealed with the development of new tests. These could perhaps include more challenging same-person trials containing stimuli found to reduce voice recognition performance (e.g., disguised vs. normal speech; Orchard & Yarmey, 1995).

## 9.1 | Dissociations between voice memory and voice matching

Hypotheses that dissociations between face memory and matching tests would transfer to the voice mode were also supported, particularly when evaluating matching tasks. As noted above, sensitivity ( $d'$ ) on the BVMT (Mühl et al., 2017) was significantly higher for the pooled Exceptional-Matcher group (SF-Identifiers, SF-Matchers), compared to the pooled Typical-Matcher group (SF-Recognisers,

Controls). The design of the BVMT (Mühl et al., 2017) is naturally different to the GFMT (Burton et al., 2010) as it accommodates audio rather than visual stimuli. However, both tasks have similar structures. With the former, participants listen to individually presented pairs of voices as many times as they like (50% matched, 50% mismatched). In the GFMT, participants can shift gaze back and forth between the simultaneously displayed faces (50% matched, 50% mismatched) for as long as they like. As such, task similarity may partly contribute to the between-group effects as well as the moderate correlations between GFMT and BVMT scores reported in Table 3. These findings are also supported by the individual data in Table 4. Compared to the typical-range face matching groups, a higher proportion of both SF-Identifiers and SF-Matchers achieved the threshold for  $>1.5$  SD super-voice processing ability on the BVMT.

In contrast to the matching task effects, SF-Recognisers, possessing exceptional face memory skills did not outperform controls on either famous or unfamiliar voice memory tests. Only when the pooled Exceptional-Recogniser group (SF-Identifiers, SF-Recognisers) was compared to the pooled typical-ability face recogniser group (SF-Matchers, Controls) were the predicted significant effects revealed, a possible consequence of greater statistical power. It is not clear why superiority at face matching (i.e., SF-Matchers, SF-Identifiers) relates to superior discrimination of matched and mismatched voices; whereas superiority at both face recognition and face matching (i.e., SF-Identifiers) is seemingly required to find effects when testing memory for voices. Indeed, many SF-Identifiers achieved the super-voice threshold on the two voice memory tests. Slightly fewer, but equally proportioned numbers of SF-Matchers and SF-Recognisers achieved criterion on the FVRT, whereas far more SF-Matchers achieved super-voice criteria than SF-Recognisers on the GVMT.

## 9.2 | Voice-specific effects

Performances on the two unfamiliar voice identity tests correlated with those on the bells section of the GVMT. However, dissociations between the patterns of results for the Bells and Voices sections of the GVMT also support the principle that, like face processing, individual differences in voice recognition may be voice-specific, in that they do not correlate with other acoustic processing tasks. No between-group effects were found for bells recognition, and bells recognition did not correlate with CFMT+ or GFMT scores. Not surprisingly therefore, the proportion of participants achieving  $>1.5$  criterion for bells was considerably lower than those achieving criterion for voices (Table 4). These voice-specific results suggest parallels with the face-specific effects shown in super-face-recognisers (e.g., Bobak, Bennets, et al., 2016; Dunn et al., 2020), as many do not display superiority on other visual processing tasks. The lack of any between-groups bells effects also suggest that the voice-specific advantages found in the current research are unlikely to be associated with motivational factors. Indeed, it would not seem credible that high performing voice identification participants would be any less motivated to attempt to accurately recognise bells. Nevertheless, these suggestions can only

be considered tentative as only bells were tested here. Future research could employ different acoustic stimuli.

Recent proposals have also suggested that a generalised inherited face-specific identity mechanism (*f*), analogous to that long-proposed for intelligence (*g*) (Spearman, 1927), might be responsible for about 25% of the variation in face identity processing ability (McCaffery et al., 2018; Verhallen et al., 2017). Although correlations and between-groups effect sizes were small in the current research, the results provide some support that this mechanism may also have a cross-modality identity-specific element. Nevertheless, further research is required to substantiate this proposal. Indeed, as stronger effects were found within the voice processing mode, it is also possible a voice-specific mechanism (perhaps 'v') may also operate.

### 9.3 | Familiar and unfamiliar voice processing

Compared to the correlations between the two unfamiliar voice tests, weaker relationships were revealed between these tests and the FVRT. This was not unexpected given that models propose different processing pathways for unfamiliar and familiar voices (Kreiman & Sidtis, 2011, 2013). The outcomes are also consistent with previous research investigating within-modality relationships between familiar and unfamiliar face tests (e.g., Lander & Poyarekar, 2015).

Intriguingly, the floor effects in the FVRT may be accounted for by test demands. As opposed to the old/new or same/different response required in the two unfamiliar voice tests, participants were required to state whether the voice was familiar or unfamiliar, and then provide identifiable, albeit brief, information about the voice heard in each trial. This was then followed by a judgement of how familiar the owner of that voice was to the participant on a 0 (not at all familiar) to 7 (extremely familiar) scale. This can be explained in terms of grain size (e.g., Yaniv & Foster, 1995), a theory of metacognition, balancing being correct in providing responses, with being helpful or informative. As such, participant may prefer to respond 'unfamiliar,' rather than providing a precise, and potentially incorrect name, despite recognising the voice in the first instance. Furthermore, it is not clear why participants generally adopted a more conservative response bias, especially since the identities of the celebrities were more or less globally known. A more liberal response would have perhaps been more expected in this circumstance. However, from some participant e-mails some non-English speakers from outside the UK and with no familiarity with the famous voices but who recognised their faces completed the tests (famous faces may be known, but their voices are often overdubbed on international TV shows) despite requests not to take part.

Mean familiarity ratings across all 28 famous trials were fairly low ( $M = 2.89$ , mid-point of scale = 3.5), which may also explain the conservative response biases.<sup>2</sup> Nevertheless, as previously discussed, and despite potential floor effects, SF-Identifiers still outperformed controls, with proportionally more achieving super-voice-recogniser status on this test, suggesting that they may develop stronger internal representations for faces (e.g., Bruce & Young, 1986;

Valentine, 1991) and voices (e.g., Kreiman & Sidtis, 2011; Lavan, Knight, & McGettigan, 2019).

#### 9.3.1 | Limitations

Although the results revealed significant effects, often supporting predictions, most generated relatively small effect sizes. This may be a consequence of the test limitations described above. Both unfamiliar voice tests contained relatively small numbers of stimuli, potentially reducing their ability to discriminate between exceptional and good performers. In addition, small effect sizes generated from comparisons between 'super' and control groups may be a consequence of controls surpassing expected population-typical skills. Despite controls only being invited if they had previously achieved typical mid-range scores on the CFMT+ and GFMT, controls outperformed a student group on the two unfamiliar voice tests—this group had no such upper restrictions on membership. As such, effect sizes in many analyses may have been weakened by a lack of controls possessing poorer face recognition abilities. Finally, the construction of all four groups were based on one face memory and one face matching test only. All research finds performance inconsistencies when examining individual differences in performances across different face processing tests purporting to measure the same factor (e.g., Bate et al., 2018, 2019). Some variation is to be expected, although poor performance on any single test may be due to multiple factors not associated with face identification ability (i.e., distractions, internet problems, motivation). This research therefore needs replication with the inclusion of far larger group samples, with face (and voice) identification ability verified using more than one test of each type.

## 10 | CONCLUSIONS

This paper presents the first research describing positive relationships between voice and face processing when examining participants with exceptional and typical-range face memory and matching ability, supporting previous research finding common cross-modal mechanisms (Tsantani et al., 2019). It was also the first research to show that a few individuals with superior face recognition ability are capable of transferring these skills to voices, raising the possibility that they may also be super-voice recognisers. Furthermore, the results demonstrated that dissociated superior performance patterns found on face recognition and face matching tests in super-recognisers extend to similar tests in the voice modality.

Finally, individuals with these skills may be able to enhance policing and security operations. Super-face-recognisers are successfully employed in specialist police units, and similar work may be available for those with exceptional voice processing skills. Algorithms can assist in the identification of faces and voices, although little research has compared human and machine performance at standardised tasks (although see Phillips et al., 2018, for evidence that super-face-recognisers achieve roughly equal accuracy to top-of-the-range face

recognition algorithms). Similarly, Hautamäki et al. (2013) found some human listeners can correctly discriminate some voices when automatic systems fail. Nevertheless, regardless of algorithm improvements, in legal contexts, critical decisions of identity will likely always be made by humans (i.e., algorithm operator, police, expert witnesses, jurors in court—see Colman, 2016 for an example), and those with superior voice processing skills may be able to perform a vital role. As such, screening for those with superior voice processing abilities might be a useful tool in the recruitment stage of certain professions.

## ACKNOWLEDGEMENT

This research was funded by the University of Greenwich Vice Chancellor's PhD Scholarship (2017/2018) awarded to the second, third, and seventh authors. Thanks go to Constanze Mühl and Virginia Aglieri for the use of the BVMT and the GVMT respectively. We also thank Rebecca Glass, Nikolay Petrov, Bethan Burnside, Katy Weatherley and Dr Elena Belanova for their assistance in data collection, inter-rater analyses and test piloting. Parts of this research have been presented at the British Society for the Psychology of Individual Differences conference, 2019, at Brunel University London; Postgraduate Researcher Conference 2019 at University of East Anglia; the British Psychological Society Cognitive Psychology Section & Developmental Psychology Section joint conference 2019, in Stoke; and the Experimental Psychology Society January London Meeting 2020.

## CONFLICT OF INTEREST

The authors declare that they had no conflicts of interest.

## ENDNOTES

<sup>1</sup> The test originally contained 40 trials; two famous voices were excluded due to experimenter error.

<sup>2</sup> Additional One-way ANOVAs on mean familiarity ratings between each group revealed a significant main effect,  $F(3, 525) = 4.91$ ,  $p = .002$ ,  $\eta_p^2 = .027$ . Post-hoc tests revealed that both SF-Identifiers and SF-Matchers responded with higher familiarity ratings than controls.

## DATA AVAILABILITY STATEMENT

All anonymised data associated with this project have been uploaded to an open access store at <https://osf.io/gqp53/>.

## ORCID

Ryan E. Jenkins  <https://orcid.org/0000-0001-7870-695X>

Claire P. Monks  <https://orcid.org/0000-0003-2638-181X>

David J. Robertson  <https://orcid.org/0000-0002-8393-951X>

Josh P. Davis  <https://orcid.org/0000-0003-0017-7159>

## REFERENCES

- Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow voice memory test: Assessing the ability to memorize and recognize unfamiliar voices. *Behavior Research Methods*, 49(1), 97–110. <https://doi.org/10.3758/s13428-015-0689-6>
- Assal, G., Aubert, C., & Buttet, J. (1981). Cerebral asymmetry and voice recognition. *Revue Neurologique*, 137(4), 255–268.
- Assal, G., Zander, E., Kremin, H., & Buttet, J. (1976). Voice discrimination in patients with cerebral cortical lesions. *Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie*, 119(2), 307–315.
- Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: A critical review and future directions. *Frontiers in Human Neuroscience*, 8(July), 1–17. <https://doi.org/10.3389/fnhum.2014.00491>
- Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., ... Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, 3(22), 1–19. <https://doi.org/10.1186/s41235-018-0116-5>
- Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Portch, E., Murray, E., & Dudfield, G. (2019). The consistency of superior face recognition skills in police officers. *Applied Cognitive Psychology*, 33(5):828–842. <https://doi.org/10.1002/acp.3525>
- Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex*, 108, 92–111. <https://doi.org/10.1016/j.cortex.2018.07.008>
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102(4), 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x>
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135. <https://doi.org/10.1016/j.tics.2004.01.008>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: A case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology*, 34(6), 357–376. <https://doi.org/10.1080/02643294.2017.1402755>
- Blank, H., Anwender, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *Journal of Neuroscience*, 31(36), 12906–12915. <https://doi.org/10.1523/jneurosci.2091-11.2011>
- Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, 82, 48–62. <https://doi.org/10.1016/j.cortex.2016.05.003>
- Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, 30(1), 81–91. <https://doi.org/10.1002/acp.3170>
- Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, 7(1378), 1–11. <https://doi.org/10.3389/fpsyg.2016.01378>
- Bredart, 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. <https://doi.org/10.1177/0963721412454876>
- Bruce, V., Henderson, Z., Newman, C., & Burton, M. A. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7(3), 207–218. <https://doi.org/10.1037/1076-898X.7.3.207>
- Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305–327. <https://doi.org/10.1111/j.2044-8295.1986.tb02199.x>
- Burton, M. A. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, 66(8), 1467–1485. <https://doi.org/10.1080/17470218.2013.800125>
- Burton, M. A., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42(1), 286–291. <https://doi.org/10.3758/BRM.42.1.286>
- Burton, M. A., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Society*, 10(3), 243–248. <https://doi.org/10.1111/1467-9280.00144>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Routledge.



- Colman, S. (2016). Voice recognition evidence: R v Kapikanya [2015] EWCA Crim 1507. *The Journal of Criminal Law*, 80(1), 5–7. <https://doi.org/10.1177/0022018315624200>
- Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience*, 10, 10–20. <https://doi.org/10.1016/j.dcn.2014.07.003>
- Davis, J. P., Forrest, C., Treml, F., & Jansari, A. (2018). Identification from CCTV: Assessing police super-recognition ability to spot faces in a crowd and susceptibility to change blindness. *Applied Cognitive Psychology*, 32(3), 337–353. <https://doi.org/10.1002/acp.3405>
- Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognitioners. *Applied Cognitive Psychology*, 30(6), 827–840. <https://doi.org/10.1002/acp.3260>
- Davis, J. P., Maguit, A., & Forrest, C. (2019). The wisdom of the crowd: A case of post- to ante-mortem face matching by police super-recognitioners. *Forensic Science International*, 302, 109910. <https://doi.org/10.1016/j.forsciint.2019.109910>
- Davis, J. P., & Robertson, D. (2020). Capitalizing on the super-recognition advantage: A powerful, but underutilized, tool for policing and national security agencies. *Journal of the Homeland Defence & Security Information Analysis Centre (HDIAC)*, 7(1), 20–25.
- Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, 23, 482–505. <https://doi.org/10.1002/acp.1490>
- Dunn, J., Summersby, S., Towler, A., Davis, J. P., & White, D. (2020). UNSW face test: A screening tool for super-reconizers. *PLoS One*, 15(11), e0241747. <https://doi.org/10.1371/journal.pone.0241747>
- Gainotti, G., Barbier, A., & Marra, C. (2003). Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain*, 126(4), 792–903. <https://doi.org/10.1093/brain/awg092>
- 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. <https://doi.org/10.1016/j.neuropsychologia.2008.08.003>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons, Ltd.
- Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia*, 48(4), 1104–1114. <https://doi.org/10.1016/j.neuropsychologia.2009.12.011>
- Hautamäki, R. G., Hautamäki, V., Rajan, P., & Kinnunen, T. (2013). *Merging human and automatic system decisions to improve speaker recognition performance*. Proceedings of the Annual Conference of the International Speech Communication Association (August), pp. 2519–2523.
- Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *The Annals of Statistics*, 12(1), 61–75. <https://doi.org/10.1214/aos/1176346392>
- Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 986–1004. <https://doi.org/10.1037/0096-1523.22.4.986>
- Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics*, 10(2), 139–148. <https://doi.org/10.1121/1.1914230>
- Jeffrey, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1824–1840. <https://doi.org/10.1037/a0025643>
- Jenkins, R., White, D., Van Montfort, X., & Burton, M. A. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323. <https://doi.org/10.1016/j.cognition.2011.08.001>
- Johnson, J., McGettigan, C., & Lavan, N. (2020). Comparing unfamiliar voice and face identity perception using identity sorting tasks. *Quarterly Journal of Experimental Psychology*, 73(10), 1537–1545. <https://doi.org/10.1177/1747021820938659>
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17(5), 577–596. <https://doi.org/10.1080/09658210902976969>
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109–2128. <https://doi.org/10.1098/rstb.2006.1934>
- Knutson, K. M., DeTucci, K. A., & Grafman, J. (2011). Implicit attitudes in prosopagnosia. *Neuropsychologia*, 49(7), 1851–1862. <https://doi.org/10.1016/j.neuropsychologia.2011.03.009>
- Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication*, 10(3), 265–275. [https://doi.org/10.1016/0167-6393\(91\)90016-M](https://doi.org/10.1016/0167-6393(91)90016-M)
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons, Ltd.
- Kreiman, J., & Sidtis, D. (2013). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Wiley-Blackwell.
- Lander, K., & Poyarekar, S. (2015). Famous face recognition, face matching, and extraversion. *Quarterly Journal of Experimental Psychology*, 68(9), 1769–1776. <https://doi.org/10.1080/17470218.2014.988737>
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2(175), 1–12. <https://doi.org/10.3389/fpsyg.2011.00175>
- Lavan, N., Burston, L. F. K., & Garrido, L. (2019). How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices. *British Journal of Psychology*, 110, 1–18. <https://doi.org/10.1111/bjop.12348>
- Lavan, N., Burston, L. F., Ladwa, P., Merriman, S. E., Knight, S., & McGettigan, C. (2019). Breaking voice identity perception: Expressive voices are more confusable for listeners. *Quarterly Journal of Experimental Psychology*, 72, 1–9. <https://doi.org/10.1177/1747021819836890>
- Lavan, N., Knight, S., & McGettigan, C. (2019). Listeners form average-based representations of individual voice identities. *Nature Communications*, 10(1), 1–9. <https://doi.org/10.1038/s41467-019-10295-w>
- Lavan, N., & McGettigan, C. (2019). *Toward a unified account of person perception from familiar and unfamiliar voices*. Paper Accessed as Pre-Print, pp. 1–24. <https://doi.org/10.31234/osf.io/shxa6>.
- Lavan, N., Merriman, S. E., Ladwa, P., Burston, L. F. K., Knight, S., & McGettigan, C. (2019). 'Please sort these voice recordings into 2 identities': Effects of task instructions on performance in voice sorting studies. *British Journal of Psychology*, 111, 1–14. <https://doi.org/10.1111/bjop.12416>
- Lavan, N., Smith, H. M. J., & McGettigan, C. (2020). Unimodal and cross-modal identity judgements using an audiovisual sorting task: Evidence for independent processing of faces and voices. Paper Accessed as Pre-Print, pp. 1–25. <https://doi.org/10.31234/osf.io/y5zbh>.
- Liu, R. R., Corrow, S. L., Pancaroglu, R., Duchaine, B., & Barton, J. J. S. (2015). The processing of voice identity in developmental prosopagnosia. *Cortex*, 71, 390–397. <https://doi.org/10.1016/j.cortex.2015.07.030>
- Macmillan, N., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press.
- McAllister, H. A., Dale, R. H. I., & Keay, C. E. (1993). Effects of lineup modality on witness credibility. *Journal of Social Psychology*, 133(3), 365–376. <https://doi.org/10.1080/00224545.1993.9712155>



- McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, M. A. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, 3(1), 21. <https://doi.org/10.1186/s41235-018-0112-9>
- McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex*, 12(1), 76–82. [https://doi.org/10.1016/S0010-9452\(76\)80033-0](https://doi.org/10.1016/S0010-9452(76)80033-0)
- Megreya, A. M., & Burton, M. A. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34(4), 865–876. <https://doi.org/10.1186/s41235-018-0112-9>
- Mühl, C., & Bestelmeyer, P. E. G. (2018). Assessing susceptibility to distraction along the vocal processing hierarchy. *Quarterly Journal of Experimental Psychology*, 72, 1657–1666. <https://doi.org/10.1177/1747021818807183>
- Mühl, C., Sheil, O., Jarutytė, L., & Bestelmeyer, P. E. G. (2017). The Bangor voice matching test: A standardized test for the assessment of voice perception ability. *Behavior Research Methods*, 50, 1–9. <https://doi.org/10.3758/s13428-017-0985-4>
- Orchard, T. L., & Yarmey, D. A. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, 9(3), 249–260. <https://doi.org/10.1002/acp.2350090306>
- Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, 85(2), 913–925. <https://doi.org/10.1121/1.397564>
- Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., Watson, R. H., Fleming, D., Crabbe, F., Valdes-Sosa, M., & Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, 119, 164–174. <https://doi.org/10.1016/j.neuroimage.2015.06.050>
- Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115(24), 6171–6176. <https://doi.org/10.1073/pnas.1721355115>
- Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, M. A. (2016). Face recognition by metropolitan police super-recognisers. *PLoS One*, 11(2), 1–9. <https://doi.org/10.1371/journal.pone.0150036>
- Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition*, 181, 12–20. <https://doi.org/10.1016/j.cognition.2018.08.004>
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252–257. <https://doi.org/10.3758/PBR.16.2.252>
- Satchell, L. P., Davis, J. P., Julle-Danière, E., Tupper, N., & Marshman, P. (2019). Recognising faces but not traits: Accurate personality judgment from faces is unrelated to superior face memory. *Journal of Research in Personality*, 79, 49–58. <https://doi.org/10.1016/j.jrp.2019.02.002>
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology European Journal of Research Methods for the Behavioural and Social Sciences*, 6(4), 147–151. <https://doi.org/10.1027/1614-2241/a000016>
- Spearman, C. (1927). *The abilities of man*. Macmillan.
- Stevenage, S. V. (2018). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia*, 116, 162–178. <https://doi.org/10.1016/j.neuropsychologia.2017.07.005>
- Stevenage, S., Symons, A., Fletcher, A., & Coen, C. (2020). Sorting through the impact of familiarity when processing vocal identity: Results from a voice sorting task. *Quarterly Journal of Experimental Psychology*, 73(4), 519–539. <https://doi.org/10.1177/1747021819888064>
- Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., ... Gosselin, F. (2019). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science*, 30(2), 300–308. <https://doi.org/10.1177/0956797618811338>
- Towler, A., Kemp, R. I., Burton, M. A., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PLoS One*, 14(2), 1–17. <https://doi.org/10.1371/journal.pone.0211037>
- Towler, A., White, D., & Kemp, R. I. (2014). Evaluating trainign methods for facial image comparison: The face shape strategy does not work. *Perception*, 43(2–3), 214–218. <https://doi.org/10.1068/p7676>
- Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, 23(1), 47–58. <https://doi.org/10.1037/xap0000108>
- Tsantani, M., Kriegeskorte, N., McGettigan, C., & Garrido, L. (2019). Faces and voices in the brain: A modality-general person-identity representation in superior temporal sulcus. *NeuroImage*, 201, 1–14. <https://doi.org/10.1016/j.neuroimage.2019.07.017>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 43(2), 161–204. <https://doi.org/10.1080/14640749108400966>
- Van Lancker, D., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, 1(2), 185–195. [https://doi.org/10.1016/0278-2626\(82\)90016-1](https://doi.org/10.1016/0278-2626(82)90016-1)
- Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, 25(5), 829–834. [https://doi.org/10.1016/0028-3932\(87\)90120-5](https://doi.org/10.1016/0028-3932(87)90120-5)
- Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, 141, 217–227. <https://doi.org/10.1016/j.visres.2016.12.014>
- Von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, 4(10), 1809–1820. <https://doi.org/10.1371/journal.pbio.0040326>
- Von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17(3), 367–376. <https://doi.org/10.1162/089929053279577>
- Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces-one element of social cognition. *Journal of Personality and Social Psychology*, 99(3), 530–548. <https://doi.org/10.1037/a0019972>
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., ... Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, 107(11), 5238–5241. <https://doi.org/10.1073/pnas.0913053107>
- Xu, X., Biederman, I., Shilowich, B. E., Herald, S. B., Amir, O., & Allen, N. E. (2015). Developmental phonagnosia: Neural correlates and a behavioral marker. *Brain and Language*, 149, 106–117. <https://doi.org/10.1016/j.bandl.2015.06.007>
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgement under uncertainty: An accuracy-informativenss trade-off. *Journal of Experimental Psychology: General*, 124(4), 424–432. <https://doi.org/10.1037/0096-3445.124.4.424>
- Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, 17(6), 263–271. <https://doi.org/10.1016/j.tics.2013.04.004>
- Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, 8, 1–9. <https://doi.org/10.3389/fnhum.2014.00562>

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, 268(1–2), 38–45. <https://doi.org/10.1016/j.heares.2010.04.011>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Jenkins RE, Tsermentseli S, Monks CP, et al. Are super-face-recognisers also super-voice-recognisers? Evidence from cross-modal identification tasks. *Appl Cognit Psychol*. 2021;35:590–605. <https://doi.org/10.1002/acp.3813>