Received: 8 November 2022 | Revised: 10 February 2023 | Accepted: 13 February 2023

DOI: 10.1002/acp.4053

RESEARCH ARTICLE

WILEY

Sounds and speech: Individual differences in unfamiliar voice recognition

Dolly Sunilkumar[1] | Steve W. Kelly[1] | Sarah V. Stevenage[2] | Dillon Rankine[1] | David J. Robertson[1]

[1]School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

[2]School of Psychology, University of Southampton, Southampton, UK

Correspondence
David J. Robertson, School of Psychological Sciences and Health, University of Strathclyde, Glasgow G1 1QE, UK.
Email: david.j.robertson@strath.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/J004995/1

Abstract

In several applied contexts (e.g., earwitness testimony), the accurate recognition of unfamiliar voices can be a critical part of the person identification process. However, recognising unfamiliar voices is prone to error. While such errors could be reduced by testing the proficiency of listeners, the established tests of unfamiliar voice matching (BVMT) and memory (GVMT) may be limited by their choice of stimuli (i.e., vowel-sounds) and their design (i.e., using identical sounds at learning and test; GVMT). Here, we examine whether these sound-based tests are predictive of performance on more naturalistic speech-based tasks, and whether performance is consistent across task-domain (matching/memory) and task-modality (voices/faces). The findings show that while the BVMT was a robust predictor of speech-based voice matching, this was not the case for the GVMT and speech-based voice memory. In addition, we provide evidence for a potential common person recognition factor '*p*'. The theoretical and applied implications are discussed.

KEYWORDS
earwitness testimony, individual differences, unfamiliar voice matching, unfamiliar voice memory, unfamiliar voice recognition

1 | INTRODUCTION

Voices convey key diagnostic cues that support identity recognition (Belin et al., 2011; Young et al., 2020). While these cues are weaker and less informative than faces (Hanley & Damjanovic, 2009; Stevenage et al., 2013), speech obtained from a criminal suspect can be submitted as forensic evidence in a court of law (Edmond et al., 2011; McGorrery & McMahon, 2017; Robson, 2017). Within the context of policing and the criminal justice system, listeners are likely to be unfamiliar with the target identity, and must rely on memory (e.g., from a crime scene interaction; Harvey et al., 2021) or perceptual matching ability (i.e., deciding whether two voices can be attributed to a common identity; Mullikin & Rahman, 2010; Smith et al., 2019, 2020), to accurately identify a voice. However, research on similar processes in unfamiliar face identification have shown that such judgements are prone to error (see Young & Burton, 2018), and several studies suggest that accurately judging the identity of unfamiliar voices might be even more problematic (see Lavan, Burton, et al., 2019).

A recent study by Kanber et al. (2022) emphasised the role of familiarity in facilitating voice identity recognition. Across two tasks, they showed that speaker identification errors were highest for unfamiliar and lab-learned voices, in comparison to voices from personally familiar individuals. This effect was present for the recognition of brief non-linguistic utterances, spoken sentences, and it remained even when the target voices had been acoustically modified. This familiarity effect, which is well documented in the face recognition literature

(see Burton et al., 2016; Johnston & Edmonds, 2009; White & Burton, 2022), has also been demonstrated in other voice recognition studies (Lavan et al., 2016; Lavan, Burston, & Garrido, 2019; Stevenage, 2018; Stevenage et al., 2020). The magnitude of unfamiliar voice recognition error represents a pressing problem in applied contexts, where misidentifications can lead to the misattribution of investigative resources and/or the conviction of an innocent suspect (Smith et al., 2019, 2020). Research has shown that the source of these errors may be a lack of awareness of the extent to which the voice of an unfamiliar person can vary (i.e., 'within-speaker' variability; see Lavan et al., 2016).

Such variation in a speaker's voice can be caused by factors such as social context, emotion, physiologic state, and speech type (Lee et al., 2019; Stevenage et al., 2021), and understanding this variance might reduce voice identification errors (see Matthews & Mondloch, 2018; Ritchie & Burton, 2017 for similiar work on faces). Lavan, Knight, et al. (2019; Experiment 3) exposed listeners to low or high variability clips of unfamiliar speakers during an identity learning phase. They reported a 5% improvement in recognition-memory accuracy for speakers that had been learned via the high (75%) compared to low (70%) variability condition. While significant, this effect is modest in size and any potential applied impact is limited by the large range of individual differences in performance (e.g., errors rates >70% for some participants). In line with the findings from Lavan, Knight, et al. (2019), attempts to improve unfamiliar face recognition have also produced modest effects that are limited by the range of individual differences in performance (e.g., Dowsett & Burton, 2015; White, Burton, et al., 2014; White, Kemp, et al., 2014).

The focus therefore has turned to the development of ecologically valid tests (see Thielgen et al., 2021) which can be used to assess identification aptitude (see Bindemann et al., 2012), and to select individuals who naturally excel at face recognition (see Bobak et al., 2016; Davis et al., 2016; Ramon, 2021; Russell et al., 2009; Wilmer et al., 2010). Here, we apply the same logic to unfamiliar voice recognition. Using tests of voice recognition ability, it should be possible to select naturally high performers for applied roles in which speaker identification is key (e.g., police investigators; Jenkins et al., 2021), as well as to assess the likely validity of voice identification judgements made by earwitnesses and jury members (Bindemann et al., 2012). There are currently two well-established tests of unfamiliar voice recognition ability: the Bangor Voice Matching Test (BVMT; Mühl et al., 2018) and the Glasgow Voice Memory Test (GVMT; Aglieri et al., 2017).

While both the BVMT and the GVMT generate a wide range of scores, they rely on tightly controlled consonant/vowel sound clips presented under ideal listening conditions (see Lavan, Burton, et al., 2019). The stimuli do not include any of the natural speaker variability (e.g., as produced during the production of full words/sentences) or environmental effects (e.g., speech recorded via phone with a noisy background environment) that listeners would encounter in regular speech. As such, it is not yet clear whether performance on these tests is indicative of more naturalistic real world speaker recognition ability. It could be the case that accurate performance on these tests simply represents a listener's ability to recognise low-level acoustic properties of isolated consent-vowel sounds (BVMT), or memory for identical vowel sound clips (GVMT). If researchers are to recommend the use of such tests to determine voice recognition aptitude, then it is critical to ensure that they do indicate likely performance in relation to real world speech (see Bate et al., 2018; Dunn et al., 2020; Thielgen et al., 2021 for similar work on faces).

Therefore, in the present study, our primary aim is to examine whether individual differences in performance on the sound-based BVMT (voice matching) and GVMT (voice memory), predict participants accuracy on speech-based voice matching (Applied Voice Matching Test; AVMaT) and memory (Applied Voice Memory Test; AVMeT) tasks. Both the AVMaT and AVMeT use full sentence voice clips which were manipulated to reflect the type of speech that listeners might encounter in the real world (e.g., speech in different locations, using different devices; Experiments 1–3). In addition, there is growing support for a general face processing factor *f*, in which face recognition aptitude is consistent across matching and memory domains (McCaffery et al., 2018; Verhallen et al., 2017; Wilmer, 2017), and here we examine whether this effect might be present for voices (Experiment 3). Finally, research has shown that there may be some degree of commonality in the processing of voices and faces (see Jenkins et al., 2021; Young et al., 2020) and so here, across the three experiments, we also include established tests of unfamiliar face matching and memory, to assess whether and to what extent identification aptitude in one modality (i.e., voices/faces) and domain (i.e., matching/memory) generalises to another (Experiments 1–3).

2 | EXPERIMENT 1

In Experiment 1, we focus on individual differences in unfamiliar voice matching. The Bangor Voice Matching Test (BVMT) consists of sound-based voice pairs (e.g., /HUD/IGI/) presented under ideal listening conditions. In contrast, the Applied Voice Matching Test (AVMaT) consists of speech-based voice pairs (i.e., full sentences) with naturalistic acoustic manipulations (e.g., speaker location, device, electronic distortion). The AVMaT uses speech-based content in the form of pre-defined sentences read by speakers in their normal/neutral tone. Using these two tests, we examine whether sound-based identification ability on the BVMT is indicative of speaker identification accuracy on the AVMaT. In addition, to assess cross-modal effects, we also include the Glasgow Face Matching Test (GFMT; Burton et al., 2010) which provides a measure of unfamiliar face matching under ideal viewing conditions, and the Models Face Matching Test (MFMT; Dowsett & Burton, 2015) which provides a more naturalistic counterpart.

3 | METHOD

3.1 | Ethics, data availability, and G*Power

Each experiment reported in this paper was approved by the Ethics Committee of the University of Strathclyde School of Psychological Sciences and Health. The data that supports the analyses reported

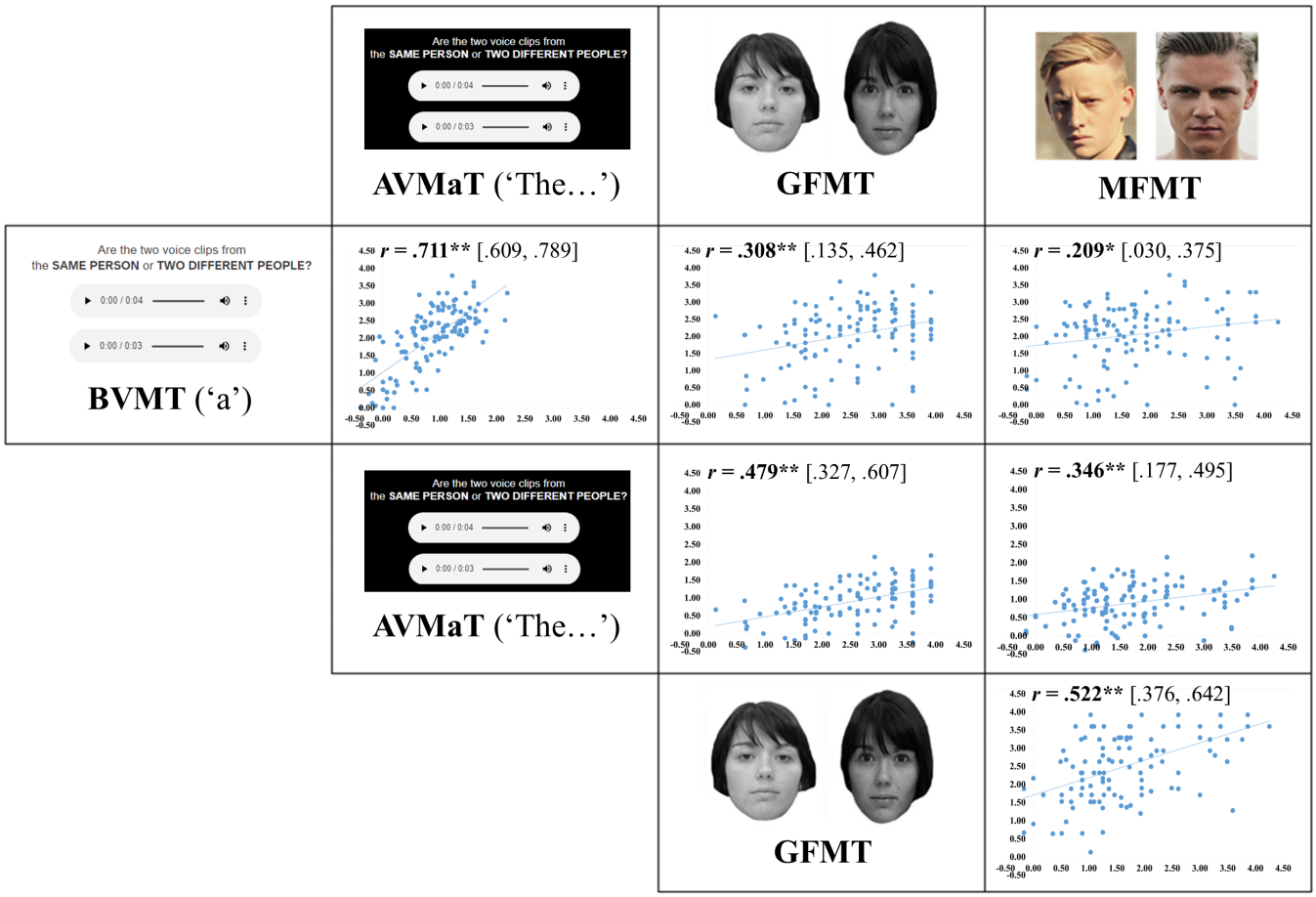

FIGURE 1 Scatterplot correlation matrix for Experiment 1. Figure 1 shows a scatterplot correlation matrix for Experiment 1 with 95% confidence intervals for the Pearson's correlation coefficients shown in square brackets. (BVMT = Bangor Voice Matching Test; AVMaT = Applied Voice Matching Test; GFMT = Glasgow Face Matching Test; MFMT = Models Face Matching Test). ** correlation is significant at the ≤ .001 level;
* correlation is significant at the < .05 level.

in this paper are available from the corresponding author upon reasonable request. G*Power analysis (Faul et al., 2009) with alpha set at .05 and power set at 80% indicated that a minimum sample of 84 participants would be required to detect a significant correlation with a medium effect size (*r* set at .3). Therefore, the sample sizes for each of the experiments reported in this paper exceed that minimum threshold to ensure that the required statistical power had been achieved ($N = 124$ for Experiment 1; $N = 95$ for Experiment 2; $N = 120$ for Experiment 3).

3.2 | Participants

A total of 124 participants with a mean age of 21 years ($SD = 5$, Range = 19–59; 85% Female) were recruited from the University of Strathclyde Psychology research participation platform. All participants reported normal or corrected to normal levels of vision and hearing, and they were granted a research participation credit on completion of the study.

3.3 | Measures

3.3.1 | Bangor voice matching test (BVMT)

The BVMT (short version) consists of 80 pairs of audio clips (40 match/40 mismatch) of consonant-vowel-consonant (e.g., /HUD/) and vowel-consonant-vowel (e.g., IGI) sounds. See Figure 1 for an example of the onscreen playback controls presented to the participants and Mühl et al. (2018) for further details.

3.3.2 | Applied voice matching test (AVMaT)

The AVMaT consists of 80 pairs of audio clips (40 match/40 mismatch) which were selected from an existing University of Southampton voice stimuli database. Speakers within this database spoke Standard Southern British English (SSBE) with a common accent and no discernible or self-reported speech impediments. All were university students aged between 18 and 30 years and all spoke English as

their first language. The test was constructed using the voices of 40 speakers (8 male) each of which was recorded reading 4 full spoken sentences. These sentences were obtained from the Home Office Centre for Applied Science and Technology FRL database, a stimulus set created for the purposes of identity recognition research (please contact the authors for further information on this), and were selected for their rich phonetic variability. They consisted of: 'the smell of freshly ground coffee never fails to entice me into the shop', 'the most important thing to remember is to keep calm and to stay safe', 'the length of her skirt caused the passers-by to stare', 'they launched into battle with all the forces they could muster'. The duration of each clip lasted between 2 s and 4 s depending on the tempo and prosody of the speaker. For the match condition, each trial was created by pairing two different clips from the same speaker. For the mismatch condition, each trial was created by pairing one clip from a speaker with a different clip from the next (same gender) speaker in the list, using a simple waterfall procedure. Each of the 4 clips were equally represented in both the match and mismatch conditions.

Pilot testing ($N = 4$) indicated that this iteration of the AVMaT, with 'clean' sound clips, would not have generated a sufficient range of scores for an individual differences analysis (i.e., accuracy rates were at or near ceiling), and so we took the opportunity to introduce environmental effects to enhance both the ecological validity and the difficulty of the task. To that end, we used the inbuilt WavePad (https://www.nch.com.au/wavepad) sound editing functions to modify the clips to sound as though the voice had been recorded in a different environment (e.g., a cave, an aircraft hangar), through different devices (e.g., telephone, CB radio), and with the addition of relevant acoustic effects (e.g., electronic noise, echo). These environments, devices, and effects were selected with applied situations in mind. For example, the cave environment was selected to emulate the type of speech encountered by the security services from terrorists operating in the middle east, who have been shown to release speech clips recorded in desert caves. Similarly, for the aircraft hangar environment, following the 9/11 terrorist attack, the prevention of aircraft hijacking has become national security priority, and we felt that this environment might reflect situations in which auditory content is obtained from suspects during their planning for such an attack. The electronic devices and distortions were also selected to try and emulate the type of real-world speech clips that practitioners are likely to encounter (i.e., captured via telephone, for example, or in situations in which the recording is of poor quality, or needs to be heard among background noise). In addition, these acoustic manipulations would also reflect situations in which police officers and security service personnel may have to identify voice content via telephones/police radio/tablets/laptops while out on investigation (i.e., not using high end sound equipment under ideal listening conditions in a quiet space).

The 80 trials were split up into 4 sets of 20 (each with 10 match/10 mismatch trials) to create a set of easy, moderate, hard, and very hard trials. For easy trials, we used the WavePad functions cave background (clip 1) and telephone + aircraft hangar (clip 2). For the moderate trials, we used the cave background + low-level acoustic distortion (clip 1) and telephone + aircraft hangar (clip 2). For the hard trials, we used the cave + low-level acoustic distortion (clip 1) and CB radio + auditorium (clip 2). Finally, for the very hard trials, we used the CB radio + low-level acoustic distortion + an auditory gargle (clip 1) and CB radio + low-level acoustic distortion + an auditory gargle + echo (clip 2). Pilot testing ($N = 10$) indicated that our trial difficulty manipulation had generated a version of the task that would be sensitive to a range of individual differences in unfamiliar voice matching performance (Overall $M = 69\%$; Range = 55%–85%; easy trials $M = 81\%$; moderate trials $M = 69\%$; hard trials $M = 65\%$; very hard trials $M = 59\%$). For the purposes of this individual differences study, a single score is calculated for each participant on the AVMaT based on response to all trials regardless of difficulty (see Duchaine & Nakayama, 2006; Russell et al., 2009).

3.3.3 | Glasgow face matching test (GFMT; short version)

The GFMT consists of 40 pairs (20 match/20 mismatch) of cropped, greyscale, forward-facing, unfamiliar faces (i.e., ideal viewing conditions). See Figure 1 for an example image pair and Burton et al. (2010) for further details.

3.3.4 | Models face matching test (MFMT)

The MFMT (short version) consists of 30 pairs (15 match/15 mismatch) of more naturalistic, unconstrained, highly variable, colour face photos of male models. See Figure 1 for an example image pair and Dowsett and Burton (2015) for further details.

3.4 | Procedure

Each experiment reported in this paper used the online testing platform Qualtrics to present the tasks and collect the data (see Germine et al., 2012). The study description informed participants that it could not be completed on a smartphone (i.e., a desktop computer/laptop/tablet was required), and they confirmed that they had followed this instruction at the start of the study. We also requested that participants complete the study in a quiet space with the use of earphones or headphones for the voice tasks. Prior to each voice task, a sound check was performed to allow listeners to adjust the volume of their device to an appropriate level. At the end of the study, participants completed a series of quality control questions which asked, 'Is there anything you would like to report to the experimenters about the study? Were you able to see and hear all the stimuli; were you able to make responses etc.?'

The order of presentation of the four tasks (BVMT, AVMaT, GFMT, MFMT) was randomised as was the trial order within each task. Following the established procedure for the BVMT (Mühl et al., 2018), the voice clips presented on each trial for the BVMT and the AVMaT could be replayed, should the participant wish to do so, until

TABLE 1 Task performance for Experiment 1.

			N = 119					
			Hits (%)			False alarms (%)		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
BVMT	Voice	Matching	86	12	38–100	23	15	0–65
AVMaT	Voice	Matching	73	14	35–98	41	13	18–75
GFMT	Face	Matching	86	16	10–100	12	12	0–45
MFMT	Face	Matching	68	18	27–100	18	15	0–73
			d Prime			Criterion *c*		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
BVMT	Voice	Matching	2.05	.87	.00–3.79	−.18	.27	−.82–.58
AVMaT	Voice	Matching	.90	.54	−.39–2.19	−.20	.29	−.98–.55
GFMT	Face	Matching	2.53	.92	.13–3.92	.03	.38	−.72–1.62
MFMT	Face	Matching	1.72	1.00	−.17–4.25	.30	.54	−1.01–1.69

Note: Mean task performance on each of the 4 identity matching tests used in Experiment 1 (BVMT = Bangor Voice Matching Test; AVMaT = Applied Voice Matching Test; GFMT = Glasgow Face Matching Test; MFMT = Models Face Matching Test).

they were confident in their match/mismatch decision. Similarly, in line with their established use, the face pairs for the GFMT/MFMT remained onscreen until response. Responses were made by clicking the onscreen label 'same person' or 'two different people' for each of the tasks, accuracy was emphasised over speed. The study took approximately 1 h, on average, to complete and this included time for participants to take screen breaks to refresh their attention.

4 | RESULTS

4.1 | Task performance

Taking a signal detection approach (Stanislaw & Todorov, 1999), we categorised a hit as a 'same person' response to a match trial and a false alarm as a 'same person' response to a mismatch trial. These values were used to calculate overall detection sensitivity (d prime; d') and response bias (c). We excluded four participants following an outlier check on d' scores (1.5*Interquartile Range; Tukey, 1977), and one participant who reported difficulty in playing the voice clips in the quality control check. Mean values and variance for each of the signal detection measures across each of the tasks are presented in Table 1. Performance across each of the established tasks was in line with published norms (Burton et al., 2010; Dowsett & Burton, 2015; Mühl et al., 2018). Importantly, the AVMaT produced a mean score which was below ceiling and above chance, and, as seen in Table 1, it produced a large range of individual differences in performance.

For the voice tasks, paired t-tests showed significantly fewer hits, t (118) = 12.20, p < .001, d = 1.12, a greater proportion of false alarms, t (118) = 16.53, p < .001, d = 1.52, and lower detection sensitivity, t (118) = 20.45, p < .001, d = 1.88, on the AVMaT compared to the BVMT. Both tasks showed a liberal response bias with no significant difference between them on this measure, t < 1. This finding shows that participants found accurate speaker identification more challenging in the speech-based AVMaT compared to the sound-based BVMT. Similarly, for the face tasks, paired t-tests showed significantly fewer hits, t (118) = 11.34, p < .001, d = 1.04, a greater proportion of false alarms, t (118) = 5.07, p < .001, d = .47, lower detection sensitivity, t (118) = 9.36, p < .001, d = .86, and a more conservative response bias, t (118) = 6.00, p < .001, d = .55, for the more naturalistic MFMT compared to the GFMT.

4.2 | Individual differences

Scatterplots, presented as a correlation matrix with Pearson's correlation coefficients (r) and 95% confidence intervals are shown in Figure 1. Importantly, as seen in Figure 1, there was a strong significant positive correlation between the sound-based BVMT and the speech-based AVMaT. This suggests that the BVMT is likely to be a robust indictor of likely performance on more naturalistic speech-based voice recognition tasks. In addition, we report significant cross-modal effects, with correlations of small-to-moderate strength between each of the voice and face tests. This finding replicates the work of Jenkins et al. (2021) for the BVMT/GFMT and extends it to include the speech based AVMaT and the more naturalistic MFMT. The correlations remained significant after applying the Benjamini-Hochberg procedure for multiple comparisons with the false discovery rate set at 10% (Benjamini & Hochberg, 1995; Jenkins et al., 2021).

5 | EXPERIMENT 2

In Experiment 2, we focus on individual differences in unfamiliar voice memory. The fallibility of eyewitness memory for faces is well

documented (Loftus, 1996; Marr et al., 2021; Wells & Olson, 2003). However, while 'earwitness' memory for voices has received much less attention in the literature, it appears to be just as fallible (see Smith et al., 2019, 2020; Stevenage et al., 2011; Yarmey, 1995). In line with Experiment 1, here we examine whether voice memory performance on an established sound-based test is predictive of speaker recognition accuracy using more naturalistic spoken sentence content. The Glasgow Voice Memory Test (GVMT) is an established measure of unfamiliar voice memory. Participants are required to learn and encode single vowel sounds (e.g., /a/) and then to recognise these same sounds from a target/foil memory test. In addition to the potential sound-based/speech-based distinction outlined in Experiment 1, this design, using identical items at learning and test, also leaves open the possibility that the GVMT may be measuring participant's memory for the stimulus rather than the speaker. Therefore, in Experiment 2, we also test participants on an Applied Voice Memory Test (AVMeT) which includes spoken sentences and, importantly, novel instances of the speaker at test. To examine cross-modal effects for memory, participants were also required to complete the Cambridge Face Memory Test (CFMT+; Duchaine & Nakayama, 2006; Russell et al., 2009).

6 | METHOD

6.1 | Participants

A total of 95 participants, who had not taken part in Experiment 1, were recruited from the University of Strathclyde Psychology research participation platform. The mean age of the sample was 22 years (SD = 6, Range = 17–51; 69% Female). All participants reported normal or corrected to normal levels of vision and hearing, and they were granted a research participation credit on completion of the study.

6.2 | Measures

6.2.1 | Glasgow voice memory test (GVMT)

The GVMT is a 16-item voice memory task in which participants are asked to recognise the voices of 8 previously learned identities that had been heard repeating the Canadian French vowel sound /a/ three times. At test, the same 8 target-identity vowel sounds that were encoded at learning are presented randomly with 8 foils. Each of the test clips are played once. As our focus is on speaker recognition rather than detecting general deficits in auditory perception, we do not include the bell learning and memory conditions of this task. See Figure 2 for an example of the onscreen playback controls presented to the participants, and Aglieri et al. (2017) for further details.

6.2.2 | Applied voice memory test (AVMeT)

The AVMeT was developed for this study, and we used the same voice identity set reported in Experiment 1 for the AVMaT. Six mismatch voice pairs were selected from the AVMaT as they produced accuracy scores that congregated around the AVMaT mean reported in Experiment 1. The rationale here was that using one identity from each pair as a target identity and the other as a foil should produce a challenging enough memory test to measure individual differences in voice recognition ability in this domain. Therefore, participants were asked to learn 6 identities (3 male/3 female) by listening to each of them repeat the sentence 'they launched into battle with all the forces they could muster' three times. At learning, the original 'clean' voice clips were used (i.e., without any auditory manipulations).

The memory test consisted of three blocks each containing 12 trials (6 target voices and 6 foils; all repeating the same sentence). Following a similar procedure to the CFMT+ (see Duchaine & Nakayama, 2006; Russell et al., 2009), in block 1, the voice clips for the 6 target identities used during the learning phase (i.e., 'they launched into battle with all the forces they could muster') were presented along with 6 foils. Importantly, in block 2, the voice clips for the 6 target identities consisted of novel instances of these speakers (i.e., using the sentence 'the smell of freshly ground coffee never fails to entice me into the shop'), with 6 foils. In block 3, the voice clips for the 6 target identities consisted of a further novel instance of the learned identities (i.e., using the sentence 'the most important thing to remember is to keep calm and stay safe'), with the 6 foils. In block 3, to add some naturalistic environmental effects, we also used the WavePad 'cave' background manipulation (i.e., the least challenging manipulation from Experiment 1), and 6 similarly adapted foils.

In line with the GVMT, participants could only hear the voice clips once during the memory block trials. Pilot testing (N = 14) indicated that the test would generate mean scores below ceiling and above chance, and importantly, a range of scores that was likely to capture individual differences in voice memory performance. In line with the AVMaT, a single score was calculated for each participant on the AVMeT based on responses to all trials regardless of difficulty (see Duchaine & Nakayama, 2006; Russell et al., 2009).

6.2.3 | Cambridge face memory test (long version; '+')

The CFMT+ is a 102-item face memory task in which participants are asked to recognise 6 previously learned identities. Each identity is learned from three example images presented in three different orientations for 3 s. Recognition memory is tested in a series of 3-AFC recognition trials that incorporate within-person variability and the addition of visual noise to increase task difficulty. See Figure 2 for a summary schematic of the CFMT+ with example stimuli, and Duchaine and Nakayama (2006) and Russell et al. (2009) for further details.

6.3 | Procedure

The procedure for Experiment 2 in relation to type of device, quiet study space, use of earphones or headphones for the voice tasks,

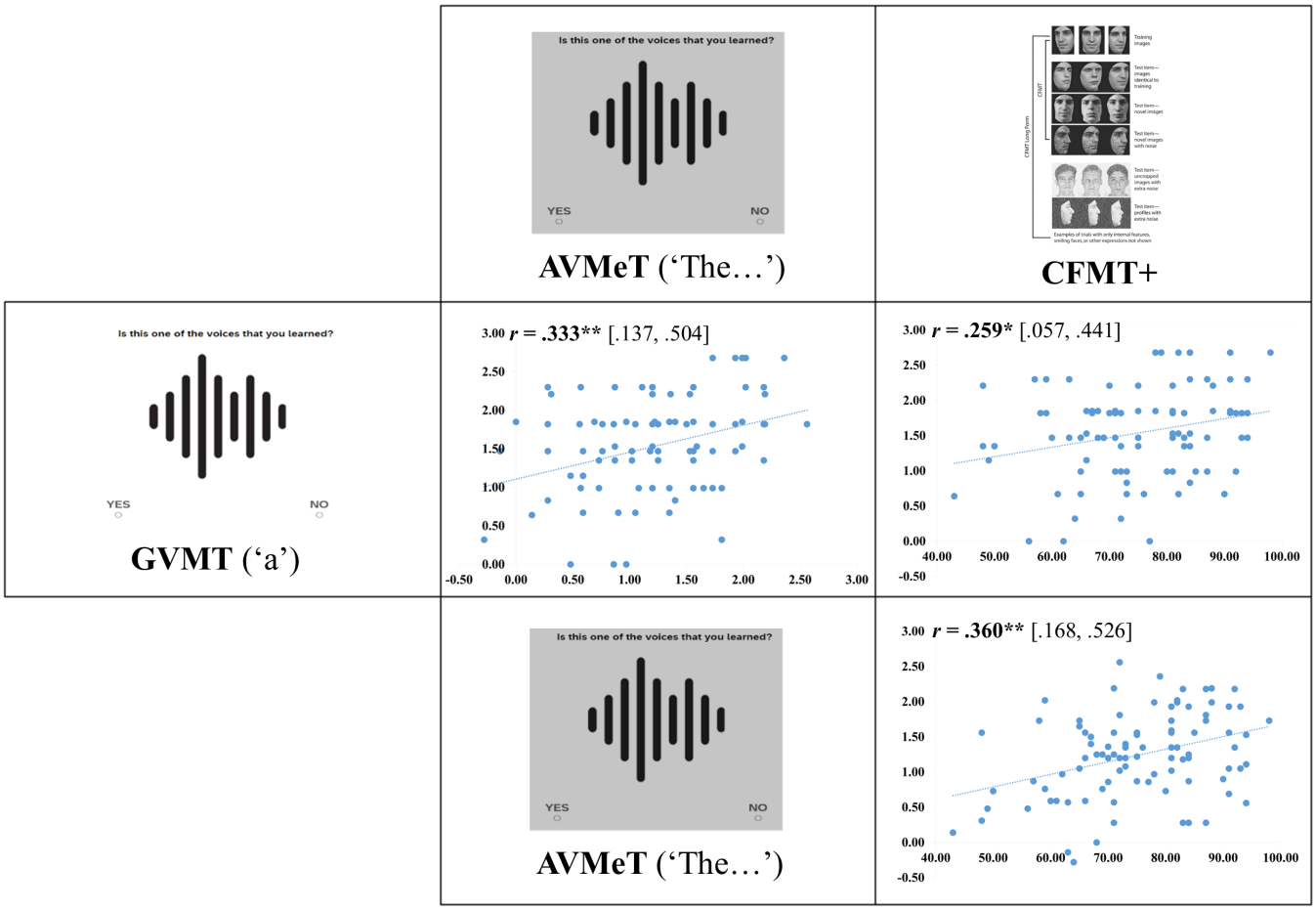

FIGURE 2 Scatterplot correlation matrix for Experiment 2. Figure 2 shows a scatterplot correlation matrix for Experiment 2 with 95% confidence intervals for the Pearson's correlation coefficients shown in square brackets. (GVMT = Glasgow Voice Memory Test; AVMeT = Applied Voice Memory Test; CFMT+ = Cambridge Face Memory Test-Long Version).
** correlation is significant at the ≤ .001 level; * correlation is significant at the < .05 level.

sound checks, and quality control questions was identical to that reported for Experiment 1. In this experiment, the order of presentation of the three tasks was fixed (GVMT, AVMeT, CFMT+) to ensure that the greater vocal content in the AVMeT did not have any carry-over effects on the GVMT. For voice memory, each clip was played to the participants once. Memory test trials for the CFMT+ remained onscreen until response. Responses were made using a clickable onscreen label (GVMT; AVMeT; Is this one of the voices you learned? 'yes', 'no'; CFMT+; Which face is one of the six target faces? '1', '2', '3'). Accuracy was emphasised over speed, with participants taking approximately 1 h, on average, with the inclusion of screen breaks, to complete the full study.

7 | RESULTS

7.1 | Task performance

For the voice memory tests, a hit was categorised as a 'yes' response to the question 'is this one of the voices that you learned' on target present trials, and a false alarm was categorised as a 'yes' response to the same question on target absent trials. These values were used to calculate overall detection sensitivity (d prime; d') and response bias (c). As the CFMT+ is a 3-AFC task, the performance measure is overall percentage accuracy (see Stanislaw & Todorov, 1999). We excluded one participant following an outlier check on d' scores, and two who reported difficulty in playing the voice clips in the quality control check. Mean values and variance for performance on each of the tasks is presented in Table 2 and scores were in line with published norms for the established tests (GVMT, CFMT+; Aglieri et al., 2017; Duchaine & Nakayama, 2006; Russell et al., 2009). Importantly, as seen in Table 2, the AVMeT produced a mean score which was below ceiling and above chance, and it detected a large range of individual differences in voice memory performance.

For the voice tasks, paired t-tests showed significantly fewer hits, t (91) = 2.72, p = .008, d = .28, a greater proportion of false alarms, t (91) = 2.98, p = .004, d = .31, and lower detection sensitivity, t (91) = 4.04, p < .001, d = .42, on the AVMeT compared to the GVMT, and both tasks showed a liberal response bias with no significant difference between them on this measure, t < 1. This finding shows that participants found accurate speaker identification more challenging in the speech-based AVMeT compared to the sound-based BVMT.

TABLE 2 Task performance for Experiment 2.

			N = 92					
			Hits (%)			False alarms (%)		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
GVMT	Voice	Memory	80	15	38–100	28	15	0–63
AVMeT	Voice	Memory	76	13	39–94	33	13	6–61
CFMT+	Face	Memory	75	12	43–98	–	–	–
			d Prime			Criterion *c*		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
GVMT	Voice	Memory	1.54	.64	.00–2.68	−.12	.36	−.77–.61
AVMeT	Voice	Memory	1.24	.61	−.28–2.56	−.14	.29	−.80–.58
CFMT+	Face	Memory	–	–	–	–	–	–

Note: Mean task performance on each of the 3 identity recognition tests used in Experiment 2 (GVMT = Glasgow Voice Memory Test; AVMeT = Applied Voice Memory Test; CFMT+ = Cambridge Face Memory Test-Long Version).

7.2 | Individual differences

Scatterplots, presented as a correlation matrix with Pearson's correlation coefficients (*r*) and 95% confidence intervals, are shown in Figure 2. Importantly, as seen in Figure 2, while there is a significant degree of correspondence in scores across the sound-based and speech-based voice memory tests, the strength of the correlation is much smaller than that reported for matching ability in Experiment 1. In other words, this finding suggest that we should place less confidence in the GVMT as robust predictor of real-world speech-based voice memory, than we should for the ability of the BVMT to predict real-world speech-based matching aptitude. This discrepancy is likely to be due to the GVMT design which relies on identical speaker stimuli at learning and test. This does not mirror the real-world unfamiliar voice recognition process, in which the listener must recognise a speaker from novel instances of their voice (e.g., from a police custody recording after having been exposed to the perpetrator's voice during a criminal act). In addition, we report significant cross-modal effects, with correlations of small-to-moderate strength between the each of the voice memory tests and the CFMT+. This finding replicates the work of Jenkins et al. (2021) for the GVMT/CFMT+ and extends it to the speech-based AVMeT. The correlations remained significant after applying the Benjamini-Hochberg procedure for multiple comparisons with the false discovery rate set at 10% (Benjamini & Hochberg, 1995; Jenkins et al., 2021).

8 | EXPERIMENT 3

The findings from Experiments 1 and 2 show that while the BVMT appears to be a robust test of naturalistic voice matching ability, this is not the case for the GVMT and voice memory ability. We also report small-to-moderate cross-modal correlations *within* each domain. Here, in Experiment 3, we investigate the potential for cross-modal *and* cross-domain effects. Previous research has shown that individuals who excel at face matching also tend to excel at face memory, this has led to the suggestion that there may be a general face processing factor '*f*', which reflects an individual's general aptitude with faces regardless of task domain (McCaffery et al., 2018; Verhallen et al., 2017). However, it is not yet well-established as to whether a similar effect, a general voice processing factor '*v*' (see Jenkins et al., 2021; Johnson et al., 2020), exists across voice tasks. In addition, while there is a growing focus on commonalities in the neural processing of voices and faces (see Young et al., 2020 for a review), few studies have assessed cross-model effects in unfamiliar voice and face identification, or indeed cross-modal and cross-domain effects (e.g., is it the case that those who excel at unfamiliar voice matching also excel at unfamiliar face memory). Therefore, in Experiment 3, participants completed the AVMaT (voice matching), the AVMeT (voice memory), the MFMT (face matching) and the CFMT+ (face memory).

9 | METHOD

9.1 | Participants

A total of 120 participants, who had not taken part in Experiments 1 or 2, were recruited from the University of Strathclyde Psychology research participation platform. The mean age of the sample was 21 years (*SD* = 4, Range = 18–43; 77% Female). All participants reported normal or corrected to normal levels of vision and hearing, and they were granted a research participation credit on completion of the study.

9.2 | Measures and procedure

For Experiment 3, we used the AVMaT (voice matching), the AVMeT (voice memory), the MFMT (face matching), the CFMT+ (face memory), and the Qualtrics online platform in an identical manner to that

TABLE 3 Task performance for Experiment 3.

			N = 116					
			Hits (%)			False alarms (%)		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
AVMaT	Voice	Matching	69	15	28–98	42	11	20–70
AVMeT	Voice	Memory	75	14	33–100	36	10	17–61
MFMT	Face	Matching	64	17	7–93	16	14	0–67
CFMT+	Face	Memory	74	14	40–99	–	–	–
			d Prime			Criterion *c*		
	Modality	Domain	*M*	*SD*	Range	*M*	*SD*	Range
AVMaT	Voice	Matching	.73	.56	−.79–2.15	−.16	.25	−.89–.44
AVMeT	Voice	Memory	1.15	.60	−.71–2.72	−.19	.26	−.87–.36
MFMT	Face	Matching	1.50	.73	−.17–3.00	.34	.36	−.97–1.63
CFMT+	Face	Memory	–	–	–	–	–	–

Note: Mean task performance on each of the 4 tests used in Experiment 3 (AVMaT = Applied Voice Matching Test; AVMeT = Applied Voice Memory Test; MFMT = Models Face Matching Test; CFMT + = Cambridge Face Memory Test-Long Version).

described in Experiments 1 and 2. The tasks were presented in a fixed order (AVMeT, CFMT+, AVMaT, MFMT), with the memory tests first followed by the matching tests, to prevent any carryover effects from exposure to the matching stimuli on the memory tests. The study took approximately 1 h and 10 min, on average, to complete.

10 | RESULTS

10.1 | Task performance

Participants' scores on the AVMaT, AVMeT, MFMT, and CFMT+ were prepared for analysis in an identical manner to that described in Experiments 1 and 2. We excluded three participants following an outlier check on *d*′ (AVMaT, AVMeT) and accuracy (CFMT+) scores, and one participant who reported difficulty in playing the voice clips in the quality control check. Task performance, presented in Table 3, was in line with published norms (CFMT+, MFMT) and in line with those reported for the AVMaT and AVMeT in Experiment 1 and Experiment 2 respectively.

10.2 | Individual differences

Scatterplots presented as a correlation matrix with Pearson's correlation coefficients (*r*) and 95% confidence intervals for each pair of tasks are shown in Figure 3. As seen in Figure 3, we replicate the cross-modal effects reported in Experiments 1 and 2. There were significant positive correlations between tests of unfamiliar voice matching (AVMaT) and unfamiliar face matching (MFMT; Experiment 1). This effect was also replicated for unfamiliar voice memory (AVMeT) and unfamiliar face memory (CFMT+; Experiment 2). We also replicate previous work which supports a general face processing factor '*f*' with a significant positive correlation, of moderate strength, between tests of unfamiliar face matching (MFMT) and unfamiliar face memory (CFMT+; Verhallen et al., 2017).

Interestingly, the novel findings from Experiment 3 show that there may also be the potential for a general voice processing factor '*v*' with a significant positive correlation between the tests of unfamiliar voice matching (AVMaT) and unfamiliar voice memory (AVMeT; Jenkins et al., 2021; Johnson et al., 2020). While this finding provides some additional support for commonalities in voice matching and memory processes, we also report significant cross-modal and cross-domain correlations between unfamiliar face memory (CFMT+) and unfamiliar voice matching (AVMaT), and between unfamiliar face matching (MFMT) and unfamiliar voice memory (AVMeT). Although these effects are small, taken together, these findings do speak to the possibility of a general person identification factor '*p*', which may reflect individual differences in identity perception aptitude regardless of modality or task domain. The correlations remained significant after applying the Benjamini-Hochberg procedure for multiple comparisons with the false discovery rate set at 10% (Benjamini & Hochberg, 1995; Jenkins et al., 2021).

11 | DISCUSSION

Person recognition errors in applied contexts can have a significant impact on the effectiveness of investigations, courtroom decision making, and the validity of criminal convictions. Minimising such errors is a major focus of applied cognitive science. Research has shown that our ability to recognise novel instances of people we are unfamiliar with is a task that is prone to error and difficult to improve (Towler et al., 2019). However, our aptitude for face and voice identification may be an innate individual difference (Aglieri et al., 2017; Shakeshaft & Plomin, 2015; Wilmer et al., 2010). We should therefore

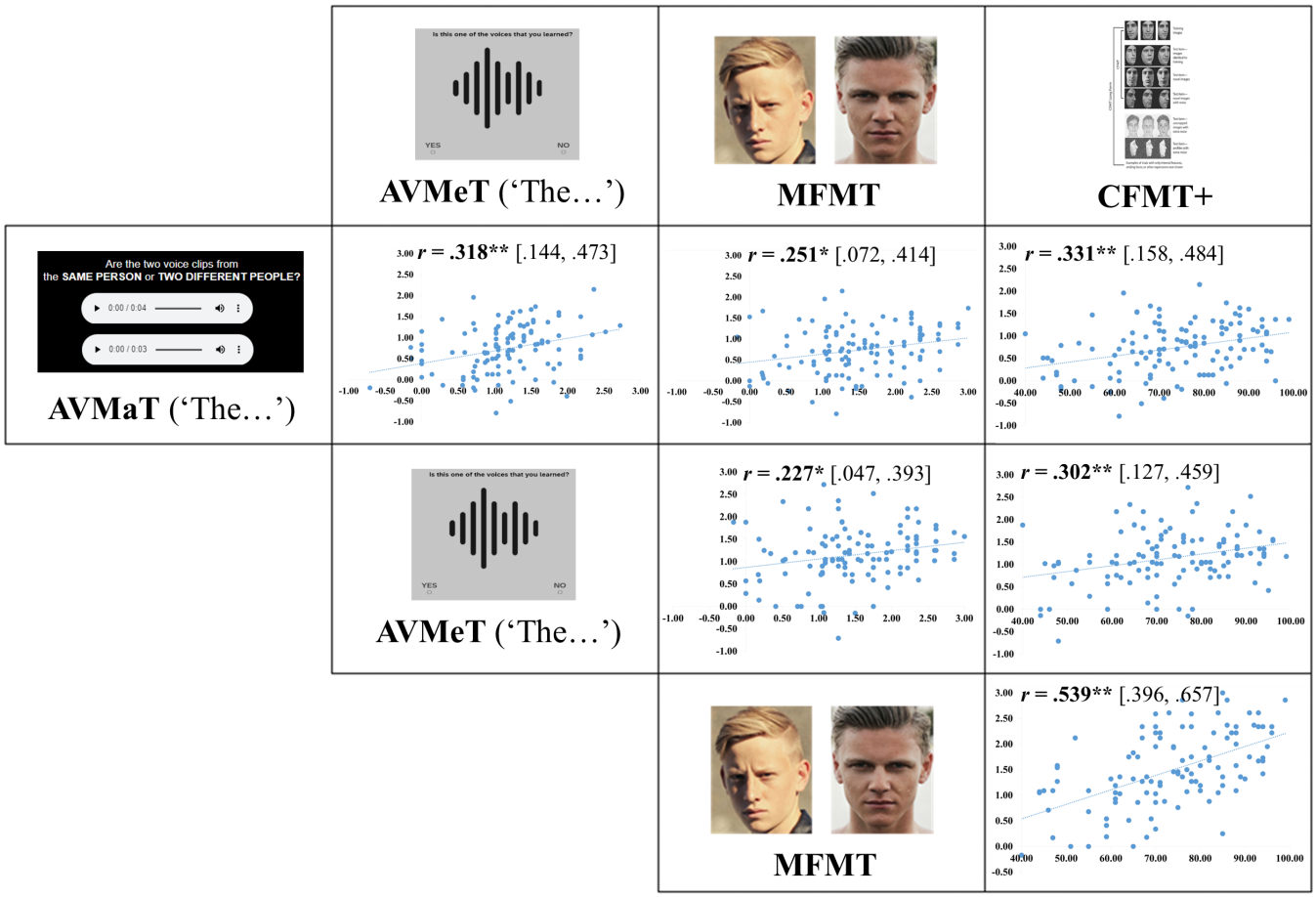

FIGURE 3 Scatterplot correlation matrix for Experiment 3. Figure 3 shows a scatterplot correlation matrix for Experiment 3 with 95% confidence intervals for the Pearson's correlation coefficients shown in square brackets (AVMaT = Applied Voice Matching Test; AVMeT = Applied Voice Memory Test; MFMT = Models Face Matching Test; CFMT+ = Cambridge Face Memory Test-Long Version). ** correlation is significant at the ≤ .001 level; * correlation is significant at the < .05 level.

be able to test the likelihood that a person will provide accurate identity judgements (e.g., eyewitness/earwitness testimony), and to select high performers for roles in which identity recognition is key (e.g., police investigators). To that end, research on faces has focused on developing ecologically valid tests of unfamiliar face recognition to ensure that lab-based tests are robust measures of likely real-world performance (see Bate et al., 2018; Dunn et al., 2020; Thielgen et al., 2021). Here we applied the same rationale and approach to unfamiliar voice recognition.

There are two established tests of ability for the recognition of unfamiliar speakers, the Bangor Voice Matching Test (BVMT; Mühl et al., 2018) and the Glasgow Voice Memory Test (GVMT; Aglieri et al., 2017). However, both tasks use isolated sound-based stimuli, presented under ideal listening conditions, and the GVMT uses the same sounds at learning and at test. Therefore, it was not clear whether aptitude as measured by these tests would predict performance on tests of more naturalistic speaker recognition, of the type that listeners would be likely to encounter in the real world. In this paper, we show that while the BVMT appears to be a robust predictor of naturalistic speech-based voice matching, this does not appear to be the case for the GVMT and naturalistic speech-based voice memory. It is likely that the GVMT is limited in this regard as it uses the same voice stimuli at learning and at test. While this is appropriate for assessing early-stage acoustic abilities, it does not appear to reflect applied processes in which speaker recognition from novel content is likely to be the task.

These findings suggest that the development of a new more ecologically valid test of unfamiliar voice memory is required for use in applied contexts. Using the AVMeT as a template, such a test should also incorporate a wider range of natural within-speaker variability, including the use of spontaneous and emotional speech (see Lavan, Burston, & Garrido, 2019), and different voice modifications to reflect more commonplace listening conditions(e.g., using a wider variety of devices and distortions, or target voices presented amidst multi-talker babble). We also endorse the approaches taken to ensure sound quality when using online testing (Woods et al., 2017) and data quality considerations articulated by Germine et al. (2012). Similarly, for unfamiliar voice matching, using the AVMaT as a template, an updated version of the test should also include greater levels of variability in speech. Future research could then examine whether the BVMT

remains a robust predictor of speaker recognition accuracy under conditions that even more closely match real-world content. The development of such tests would then support two steps that could minimise voice identification errors in applied contexts.

First, these tests could be used in support of the proposal by Bindemann et al. (2012) to examine the identity recognition ability of eyewitnesses/earwitnesses and jury members. In doing so, a more objective level of evidential weight could be attributed to their identification decisions (i.e., how likely is it that the observer/listener is making the correct identity judgement). Second, these tests could be used to select individuals who would appear to naturally excel at unfamiliar voice recognition, for roles in which speaker identification is key (see Hollien, 2002). Within the face literature, such individuals are called 'super-face-recognisers', and recent work suggests that there might also be 'super-voice-recognisers' (see Aglieri et al., 2017; Jenkins et al., 2021; and Bobak et al., 2016 for discussion on divergence in identification abilities within an identification modality). In lieu of effective training methods to improve unfamiliar identity recognition (Lavan, Knight, et al., 2019; Towler et al., 2019), the selection of such individuals, perhaps paired with our most effective algorithms (Phillips et al., 2018), might be the best current route to reducing voice recognition errors in investigative and forensic contexts.

From a theoretical perspective, recent work has suggested that the correspondence in levels of aptitude on tests of unfamiliar face matching and memory might be explained by a general face processing factor *f* (McCaffery et al., 2018; Verhallen et al., 2017). Here, with significant correlations between the voice matching and memory tests (AVMaT/AVMeT; Experiment 3), we provide evidence for a similar effect, or general voice processing factor '*v*', which might support cross-domain individual differences in performance. However, we also report small cross-domain (i.e., voice/face) and cross-modal (i.e., matching/memory) effects. This replicates and extends work from Jenkins et al. (2021) and Johnson et al. (2020), and leaves open the possibility for a common, modality-general, mechanism (a person identification factor '*p*') that underpins performance for faces and voices alike. While any such mechanism would likely play a small role in the overall person identification process, there do appear to be areas of the brain that support the multimodal processing of face and voice signals at early stages in the identity recognition process (see Young et al., 2020 for a review). It could therefore be the case that individual differences in the efficacy of those cortical regions might give rise to the cross-modal/cross-domain individual differences reported here.

To conclude, across three experiments, we show that while the sound-based BVMT appears to be a robust predictor of more naturalistic speaker recognition, this was not the case for the GVMT. We recommend the development of more ecologically valid tests of unfamiliar voice matching and memory. Our findings also support further work into the common processes which might underlie individual differences in identity recognition aptitude regardless of modality and domain. Taken together, our findings suggest a new route, based on an individual differences approach, to minimise the impact of unfamiliar voice identification errors in applied contexts.

ACKNOWLEDGMENTS

We thank Hannah Wilson for her contribution to the study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the analysis reported in this paper is available from the corresponding author upon reasonable request.

ORCID

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

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

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(1), 1–19. https://doi.org/10.1186/s41235-018-0116-5

Belin, P., Bestelmeyer, P. E., 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

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

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96–103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bobak, A. K., Hancock, P. J., & 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

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202–223. https://doi.org/10.1111/cogs.12231

Burton, A. M., 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

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology*, *30*(6), 827–840. https://doi.org/10.1002/acp.3260

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, *106*(3), 433–445. https://doi.org/10.1111/bjop.12103

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Dunn, J. D., Summersby, S., Towler, A., Davis, J. P., & White, D. (2020). UNSW Face Test: A screening tool for super-recognizers. *PLoS One*, *15*(11), e0241747. https://doi.org/10.1371/journal.pone.0241747

Edmond, G., Martire, K., & Roque, M. S. (2011). 'Mere guesswork': Cross-lingual voice comparisons and the jury. *Sydney Law Review*, *33*(3), 395–425. https://search.informit.org/doi/abs/10.3316/agispt.20115174

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

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*(8), 830–839. https://doi.org/10.1080/09658210903264175

Harvey, M. B., Bruer, K. C., & Price, H. L. (2021). Perceptions of familiar and unfamiliar ear-and eyewitnesses. *Psychiatry, Psychology and Law*, *29*(3), 395–412. https://doi.org/10.1080/13218719.2021.1910588

Hollien, H. F. (2002). *Forensic voice identification*. Academic Press.

Jenkins, R. E., Tsermentseli, S., Monks, C. P., Robertson, D. J., Stevenage, S. V., Symons, A. E., & Davis, J. P. (2021). Are super-face-recognisers also super-voice-recognisers? Evidence from cross-modal identification tasks. *Applied Cognitive Psychology*, *35*(3), 590–605. https://doi.org/10.1002/acp.3813

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

Kanber, E., Lavan, N., & McGettigan, C. (2022). Highly accurate and robust identity perception from personally familiar voices. *Journal of Experimental Psychology: General*, *151*(4), 897–911. https://doi.org/10.1037/xge0001112

Lavan, N., Burston, L. F., & Garrido, L. (2019). How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices. *British Journal of Psychology*, *110*(3), 576–593. https://doi.org/10.1111/bjop.12348

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lavan, N., Knight, S., Hazan, V., & McGettigan, C. (2019). The effects of high variability training on voice identity learning. *Cognition*, *193*, 104026. https://doi.org/10.1016/j.cognition.2019.104026

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Impaired generalization of speaker identity in the perception of familiar and unfamiliar voices. *Journal of Experimental Psychology: General*, *145*(12), 1604–1614. https://doi.org/10.1037/xge0000223

Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *The Journal of the Acoustical Society of America*, *146*(3), 1568–1579. https://doi.org/10.1121/1.5125134

Loftus, E. F. (1996). *Eyewitness testimony*. Harvard University Press.

Marr, C., Sauerland, M., Otgaar, H., Quaedflieg, C. W., & Hope, L. (2021). The effects of acute stress on eyewitness memory: An integrative review for eyewitness researchers. *Memory*, *29*(8), 1091–1100. https://doi.org/10.1080/09658211.2021.1955935

Matthews, C. M., & Mondloch, C. J. (2018). Improving identity matching of newly encountered faces: Effects of multi-image training. *Journal of Applied Research in Memory and Cognition*, *7*(2), 280–290. https://doi.org/10.1016/j.jarmac.2017.10.005

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1), 1–15. https://doi.org/10.1186/s41235-018-0112-9

McGorrery, P. G., & McMahon, M. (2017). A fair 'hearing' earwitness identifications and voice identification parades. *The International Journal of Evidence & Proof*, *21*(3), 262–286. https://doi.org/10.1177/1365712717690753

Mühl, C., Sheil, O., Jarutytė, L., & Bestelmeyer, P. E. (2018). The Bangor Voice Matching Test: A standardized test for the assessment of voice perception ability. *Behavior Research Methods*, *50*(6), 2184–2192. https://doi.org/10.3758/s13428-017-0985-4

Mullikin, A., & Rahman, S. S. (2010). The ethical dilemma of the USA government wiretapping. *International Journal of Managing Information Technology*, *2*(4) https://ssrn.com/abstract=3393479, 32–39.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C. D., Chellappa, R., White, D., & 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

Ramon, M. (2021). Super-recognizers–a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, *158*, 107809. https://doi.org/10.1016/j.neuropsychologia.2021.107809

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656

Robson, J. (2017). A fair hearing? The use of voice identification parades in criminal investigations in England and Wales. *Criminal Law Review*, *1*, 36–50 http://irep.ntu.ac.uk/id/eprint/29636

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

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, *112*(41), 12887–12892. https://doi.org/10.1073/pnas.1421881112

Smith, H. M., Baguley, T. S., Robson, J., Dunn, A. K., & Stacey, P. C. (2019). Forensic voice discrimination by lay listeners: The effect of speech type and background noise on performance. *Applied Cognitive Psychology*, *33*(2), 272–287. https://doi.org/10.1002/acp.3478

Smith, H. M., Bird, K., Roeser, J., Robson, J., Braber, N., Wright, D., & Stacey, P. C. (2020). Voice parade procedures: Optimising witness performance. *Memory*, *28*(1), 2–17. https://doi.org/10.1080/09658211.2019.1673427

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

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. V., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology*, *25*(1), 112–118. https://doi.org/10.1002/acp.1649

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. https://doi.org/10.1007/s00426-012-0450-z

Stevenage, S. V., Symons, A. E., 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), 519–536. https://doi.org/10.1177/1747021819888064

Stevenage, S. V., Tomlin, R., Neil, G. J., & Symons, A. E. (2021). May I speak freely? The difficulty in vocal identity processing across free and scripted speech. *Journal of Nonverbal Behavior*, *45*(1), 149–163. https://doi.org/10.1007/s10919-020-00348-w

Thielgen, M. M., Schade, S., & Bosé, C. (2021). Face processing in police service: The relationship between laboratory-based assessment of face processing abilities and performance in a real-world identity matching task. *Cognitive Research: Principles and Implications*, *6*(1), 1–8. https://doi.org/10.1186/s41235-021-00317-x

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PLoS One*, *14*(2), e0211037. https://doi.org/10.1371/journal.pone.0211037

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing.

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

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*(1), 277–295. https://doi.org/10.1146/annurev.psych.54.101601.145028

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, *1*(5), 287–300. https://doi.org/10.1038/s44159-022-00041-3

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, *20*(2), 166–173. https://doi.org/10.1037/xap0000009

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100–106. https://doi.org/10.3758/s13423-013-0475-3

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & 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

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, *1*(4), 792–816. https://doi.org/10.1037/1076-8971.1.4.792

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences*, *22*(2), 100–110. https://doi.org/10.1016/j.tics.2017.11.007

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

How to cite this article: Sunilkumar, D., Kelly, S. W., Stevenage, S. V., Rankine, D., & Robertson, D. J. (2023). Sounds and speech: Individual differences in unfamiliar voice recognition. *Applied Cognitive Psychology*, 1–13. https://doi.org/10.1002/acp.4053