

IAFPA 2021



**An investigation of
the effects of
voice sample duration
and number of foils on
voice parade performance
across accents**

Kirsty McDougall, Nikolas Pautz,
Harriet M.J. Smith,
Katrin Müller-Johnson, Alice Paver
and Francis Nolan



IVIP: Improving Voice Identification Procedures



Project Team

University of Cambridge:

PI: Kirsty McDougall (phonetics)
Francis Nolan (phonetics)
Alice Paver (phonetics)

Nottingham Trent University:

Harriet Smith (psychology)
Natalie Braber (sociolinguistics)
David Wright (sociolinguistics)

De Montfort University:

Nikolas Pautz (psychology)

University of Oxford:

Jeremy Robson (law)

Katrin Müller-Johnson
(criminology, psychology)

UK ESRC Grant Ref: ES/S015965/1



Improving Voice Identification Procedures (IVIP)



- 4 different strands:

Strand 1: What are the optimal parameter values for voice parade procedures?

Strand 2: What are the psycho-phonetic underpinnings of voice distinctiveness?

Strand 3: How do social stereotypes affect voice identification?

Strand 4: How accurate are the normative assumptions of criminal justice practitioners in respect of voice identification procedures?

Outline



- Background – need for more research into the role of system variables in voice parades
- Experiment 1 – parade sample duration
- Experiment 2 – parade size
- Discussion and implications

Voice parade procedure in England and Wales



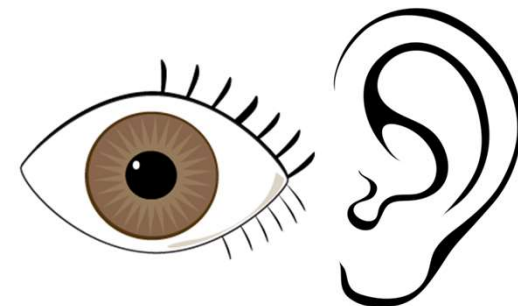
- Voice parade guidelines published in 2003 Home Office Circular
- Developed by DS (now DCI) John McFarlane and Prof. Francis Nolan in conjunction with 2001 arson case



Visual versus auditory



- Procedural aspects of VP method were based on the existing protocol for visual parades
- Nolan was encouraged at the time by the police to devise the procedure this way to reduce the chance of challenge
- Yet eyes and ears are not the same
 - visual and auditory memory operate in different ways
(Belin et al. 2004, Stevenage et al. 2011, 2012, Smith et al. 2016)
- Have the optimal parade parameters been chosen in the VP method?
 - Earwitness voice recognition is under-researched, esp. re system variables
 - System variable choices should support optimal earwitness performance



Home Office guidelines (2003): Considering some of the parameters



- Voice parade = 9 voices

- Voice samples should be 1 minute long

- Witness must be instructed that the voice of the suspect may or may not be present

- Witness is asked for a decision after listening to all voices (rather than yes/no after each voice)

- Witness is allowed to listen to the samples as many times as they wish

Does a 9-sample parade afford optimal recognition?

(cf. Bull & Clifford 1999, Levi 1998)

Parade will contain 9 mins material.
Does this lead to optimal recognition?

(cf. Smith et al. 2020)

Could the way that this is worded alter the rate of false alarms?

Does this serial format lead to optimal recognition?

(cf. Seale Carlisle & Mickes, 2016; Smith et al., 2020)

Does this lead to optimal recognition?

(cf. Pozzulo and Lindsay 1999 re elimination lineups)

Could interference be at play?

(cf. Stevenage, Howland and Tippelt 2011)

Experiment 1 - sample duration



Can sample durations be reduced without a performance cost?

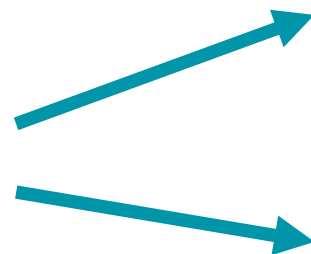
- Parade samples taken from police interview recordings, with excerpts of the suspect's voice spliced together
 - content shouldn't convey identifying information, or information relating to the crime
 - same process for foils
- Constructing nine 60s voice samples of similar-sounding people very time-consuming
- Crucially may increase the length of time between crime and parade, thus compromising memory

Experiment 1 - sample duration



Can sample durations be reduced without a performance cost?

Voice
samples
should be
60s long



People can extract basic identity information from much shorter durations

(Bestelmeyer et al., 2010; McAleer et al., 2014)

Temporal ratio models of memory - possible that shorter sample durations may lead to reduced interference between the stimuli

(Bjork and Whitten, 1974; Brown et al., 2007)

- Comparison of recommended 60s sample with 15s and 30s
- Hypothesis: shorter sample duration times would either be better, or at least no worse than the 60s sample condition.

Experiment 1 design



- 3 (sample duration) x 2 (target present/absent) between-subjects design
- Sample durations: 15s, 30s, 60s
- 6 target speakers

Speech material



Accent	Database	Sex and age	No. speakers	No. targets	No. possible foils selected	No. foils resulting
SSBE	DyViS	male, 18-25	100	3	45	27
York English	YorViS	male, 18-25	21	1	15	9
Bradford English	WYRED	male, 18-30	60	1	15	9
Wakefield English	WYRED	male, 18-30	60	1	15	9

Same speaking tasks: mock police interview; telephone call

DyViS: Nolan et al. (2009);

YorViS: McDougall et al. (2015);

WYRED: Gold et al. (2018)

Exposure and parade materials



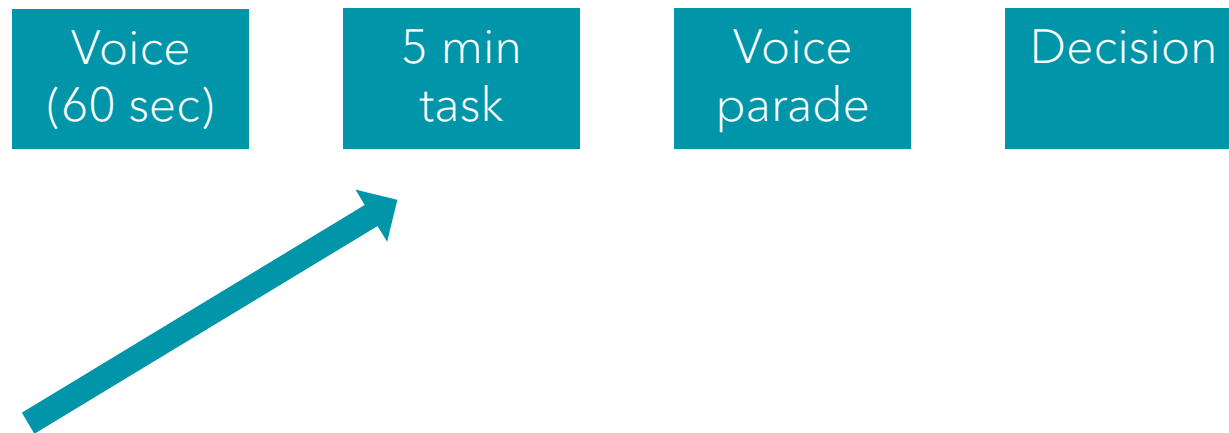
- Exposure material: 60s sample from telephone call (target side)
- Foils chosen on the basis of MDS experiment already described (McDougall et al. 2021 IAFPA):
 - 9 (for target-absent) or 8 (for target-present) speakers most similar-sounding to the target
- Parade samples: 15s, 30s, 60s samples from simulated police interview task, using collage technique of Home Office guidelines
- Experiment conducted online using Gorilla

Participants



- N = 271 participants recruited via Prolific (45 per target speaker)
 - born in and lived most of their pre-18 lives in England
 - 1st language English
 - No hearing loss or hearing difficulties
 - 136 male, 135 female, aged 18-40 years (M= 27.68, SD = 6.1)

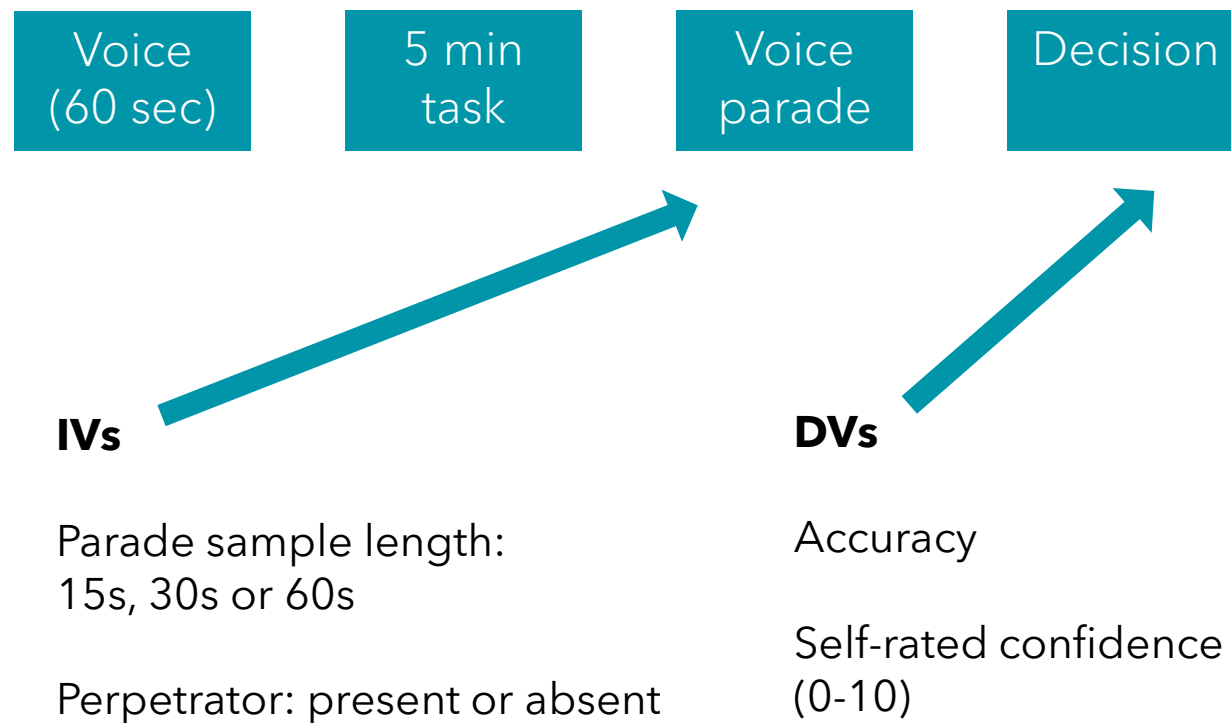
Procedure



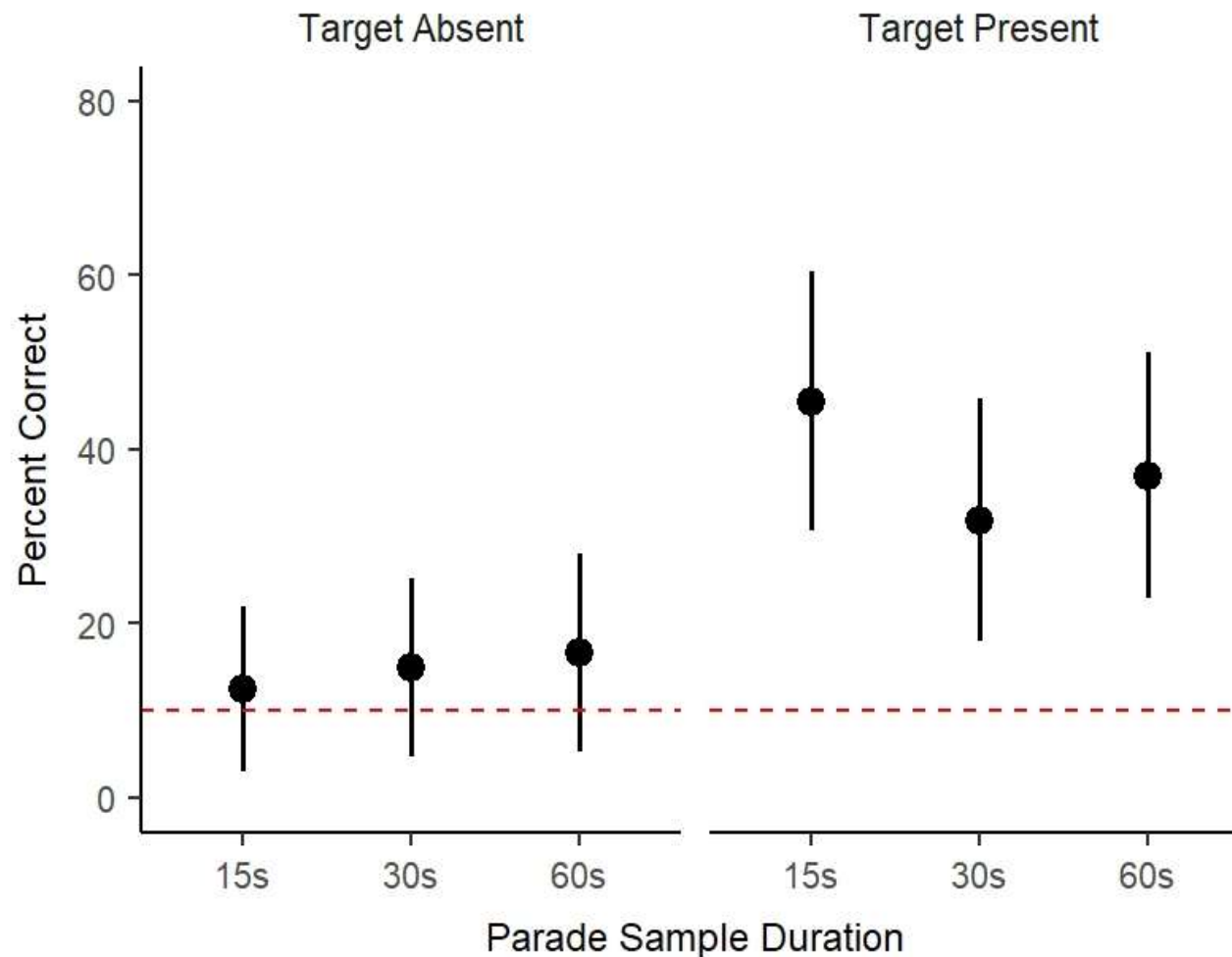
Retention interval:

- 5 min distractor task
(word search, accompanied by lobby noise)
- exceeds short-term memory capacity; relies on long-term-memory

Procedure

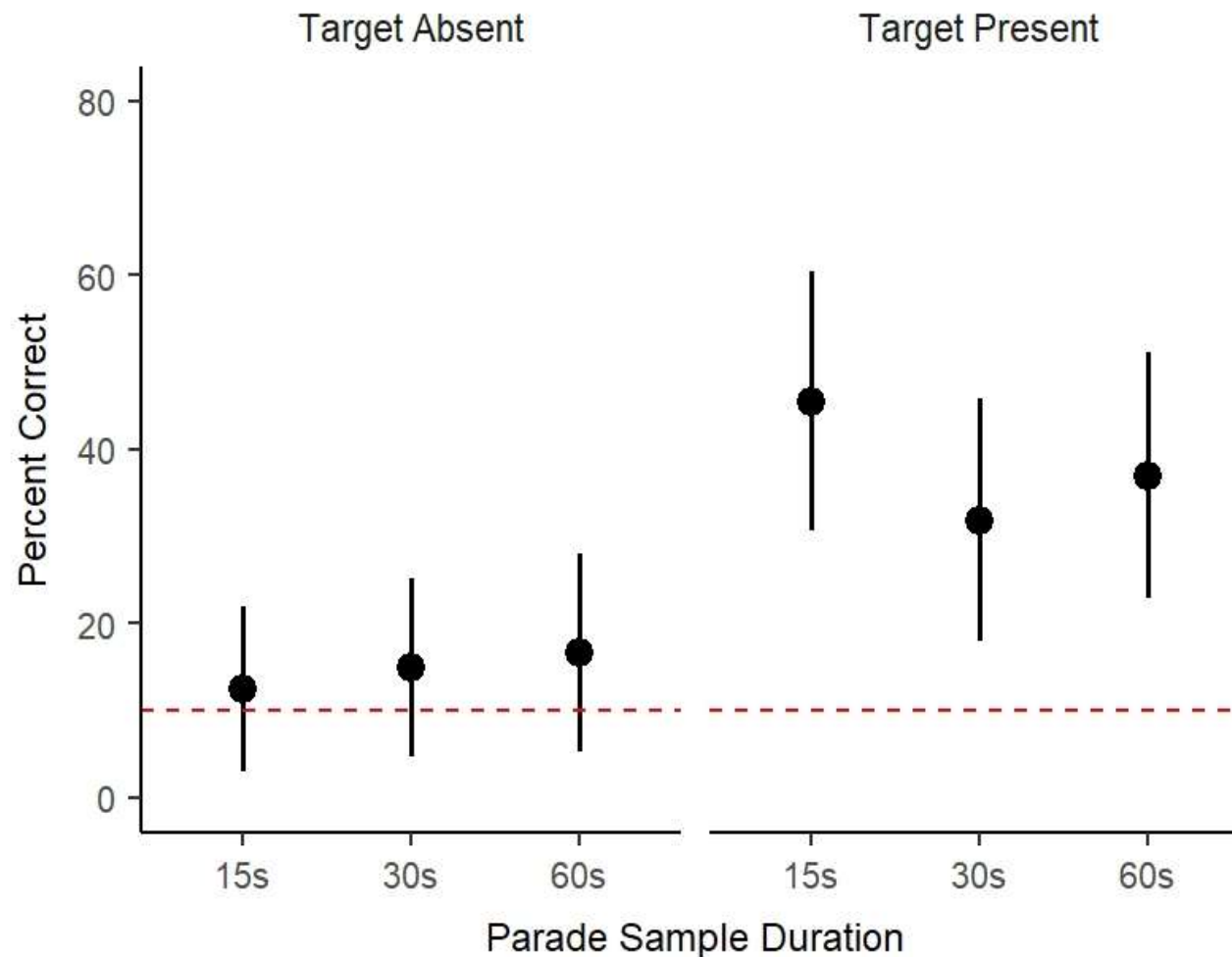


Results: effect of sample length on accuracy



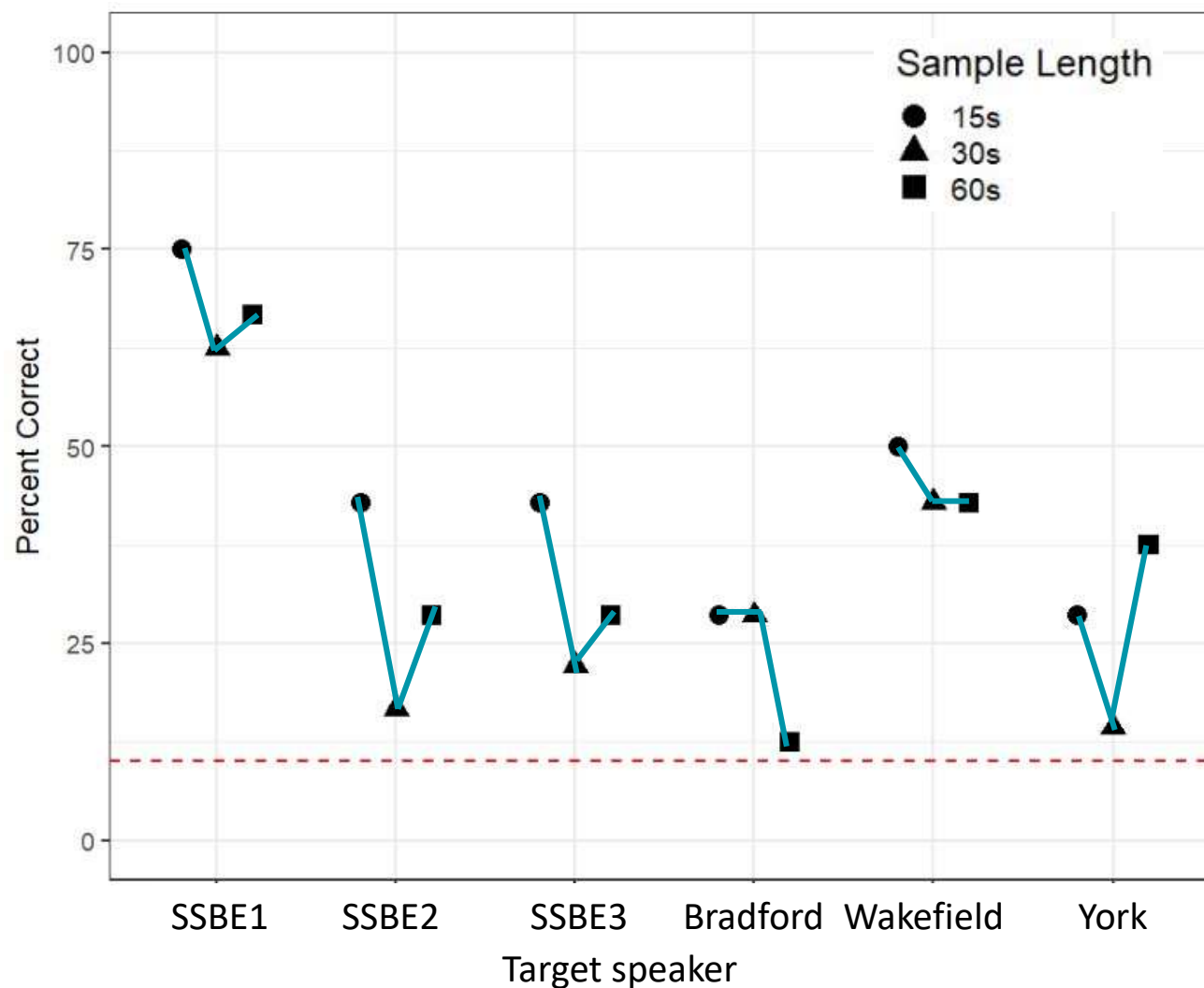
- Hit rate is relatively low, correct rejections in target-absent low
- Data were analysed using Bayesian mixed models and Signal Detection Theory analysis
- Main effect of Target Presence, with higher accuracy for target-present vs target-absent

Results: effect of sample length on accuracy



- For target-present parades, 15s samples give best performance (45% correct)
- 38% accurate for 60s samples (as per current procedure), yet slight dip to 36% for 30s
- No statistically significant difference between sample durations
- Suggests sample duration could be safely reduced for parades

Results: target-present, by target speaker



- Descriptive pattern of 15s better than 30s present for all speakers but Bradford
- 15s mostly better than 60s
- Substantially different accuracy rates for individual target speakers

Experiment 1 - Discussion



- No significant differences between 15s, 30s and 60s samples
- Suggests Home Office procedure could be satisfactorily modified by reducing sample duration to between 15 and 30s
 - Substantial reduction in preparation time for phonetician
 - May increase number of candidate foil recordings available
- Large variation in recognizability of target speakers
 - importance of including multiple targets in experiments

Experiment 2 - Parade Size



Can parade size be reduced without a performance cost?

Voice parades
should consist of
9 voices

Practical considerations

Larger lineups offer more protection to innocent suspect?

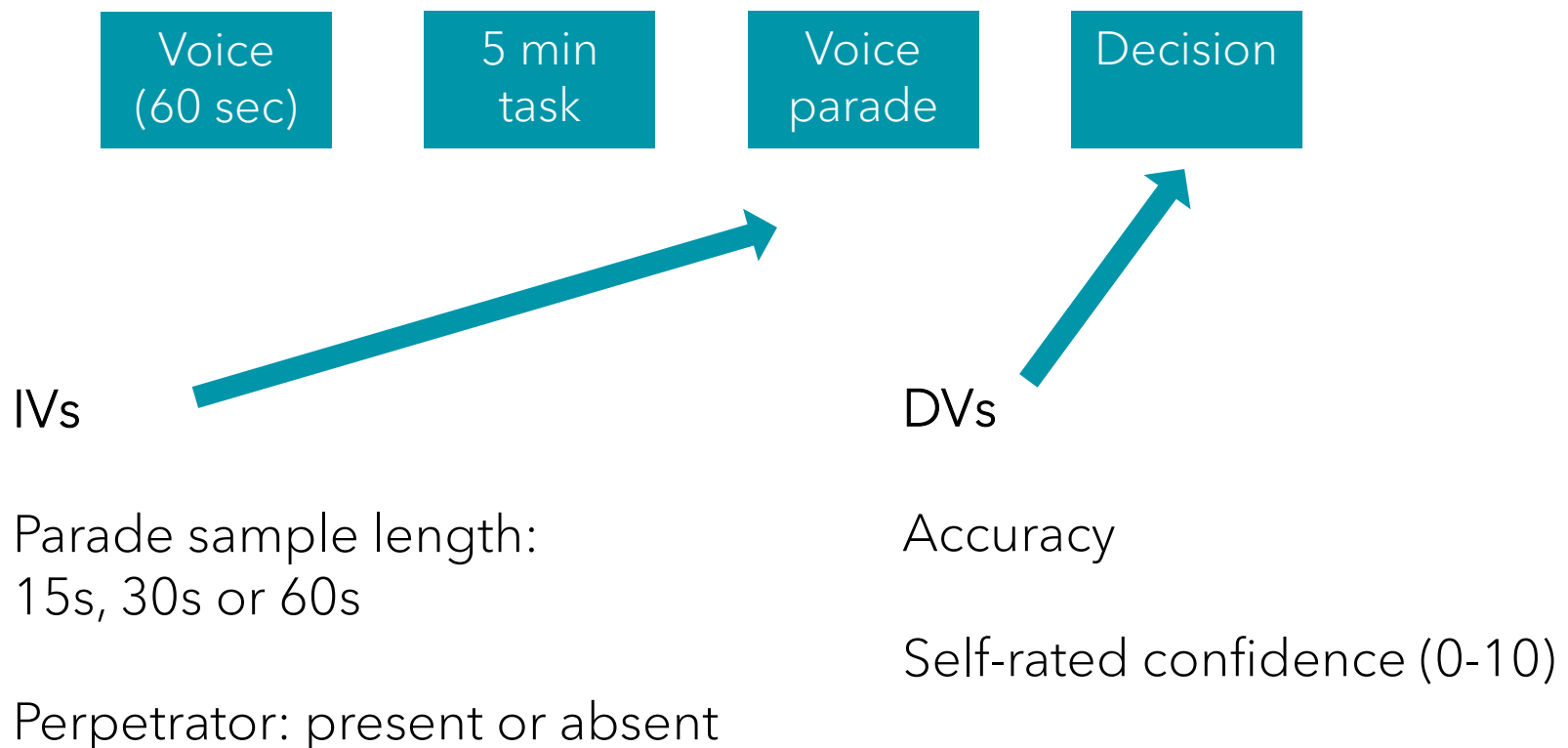
But erroneous auditory attention is more likely to occur when the demand on resources is high

(Zimmerman, Moscovitch & Alain, 2016)

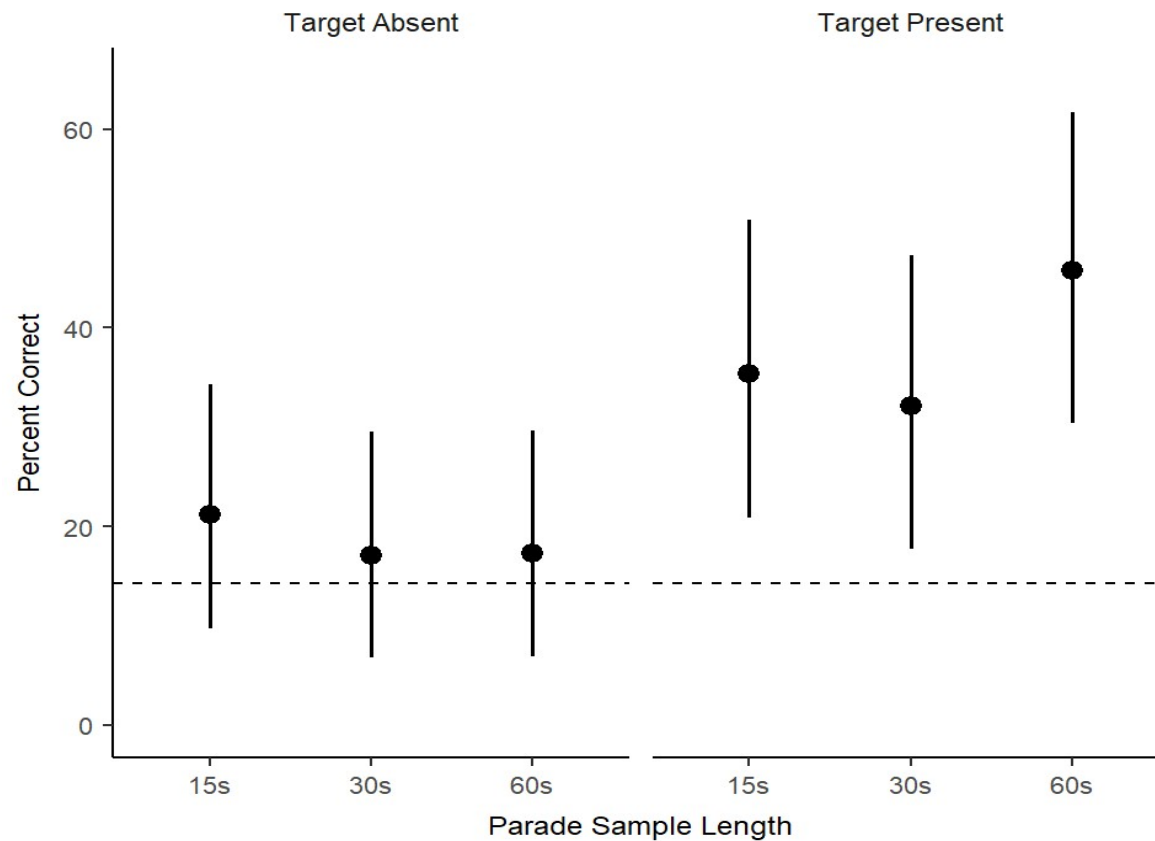
Procedure



- Same structure as Experiment 1; parades with 6 voices instead of 9
- Participants: $N=278$ (136 female)



Accuracy



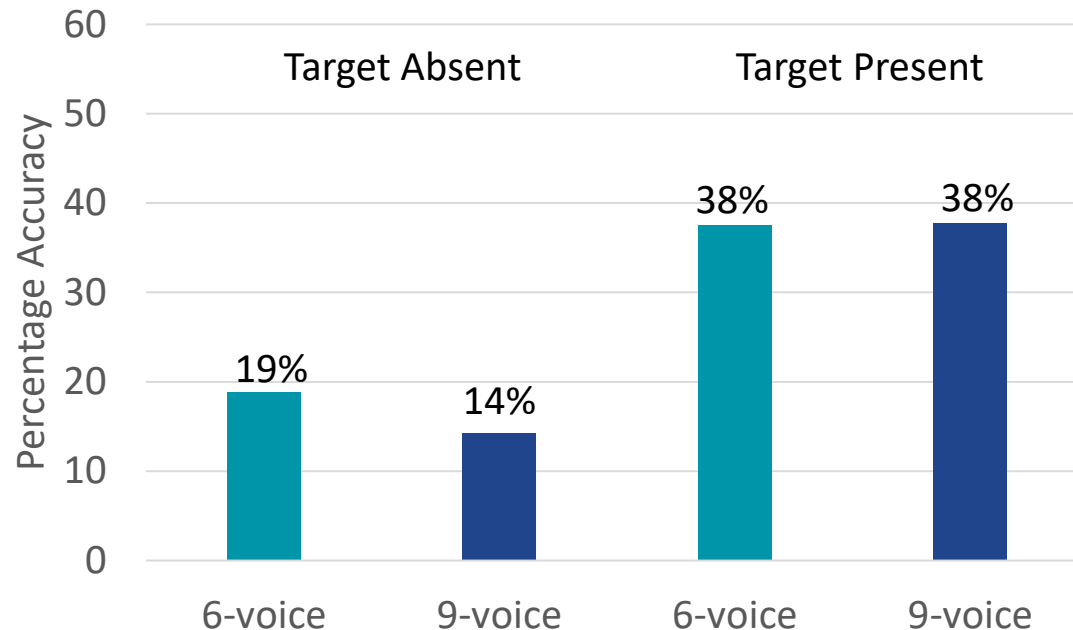
- No effect of sample duration, as in Experiment 1
- Only an effect of parade type - accuracy higher on target-present parades

Experiment 1 and 2 comparison



- No meaningful differences in accuracy between sample durations in either experiment
- Collapsed the data across duration conditions to investigate differences between the 6- and 9-person voice parades....

Experiment 1 and 2 comparison





Parade size: = 0.39 [95% HDPI: -0.29, 1], BF = 0.183;

Parade Size x Target Presence: = -0.36 [95% HDPI: -1.14, 1], BF = 0.192

- Target Absent: accuracy descriptively higher in 6- vs 9-voice
- Target Present: accuracy almost identical between 6- and 9-voice
- Bayesian mixed model results:
 - Parade Size NS; Interaction Parade Size x Target Presence NS
- 6-person parade did not improve performance

Conclusions



- Poor performance – but the task is not impossible if the target is present
 - Identification rates here probably more conservative than real-world situation
 - Very difficult task – short exposure, incidental memory (rather than intentional), complicated distractor task
- Very different performance for different target speakers
- Reduce sample duration? 
 - Home Office procedure could be satisfactorily modified by reducing sample duration from 60s
- Reduce the number of foils? 
 - However strong the warnings given, some earwitnesses will be inclined to guess and make a false identification when the target speaker is absent
 - Larger parade size affords better statistical protection of innocent suspect

See IVIP website for updates



<https://www.phonetics.mml.cam.ac.uk/ivip/>



Economic
and Social
Research Council

References 1



- Belin, P., S. Fecteau, and C. Bedard. 2004. 'Thinking the voice: neural correlates of voice perception', *Trends in Cognitive Sciences*, 8: 129-35.
- Bestelmeyer, P.E.G., J. Rouger, L. M. DeBruine, and P. Belin. 2010. 'Auditory adaptation in vocal affect perception', *Cognition*, 117: 217-23.
- Brown, G.D.A., I. Neath, and N. Chater. 2007. 'A temporal ratio model of memory', *Psychological Review*, 114: 539-76.
- Bjork, R.A., and W.B. Whitten. 1974. 'Recency-sensitive retrieval processes in long-term free recall', *Cognitive Psychology*, 6: 173-89.
- Bull, R., and B. Clifford. 1999. 'Earwitness testimony.' in Anthony Heaton-Armstrong, Eric Shepherd and David Wolchover (eds.), *Analysing Witness Testimony: A Guide for Legal Practitioners and Other Professionals* (Blackstone: London).
- Gold, E., S. Ross, and K. Earnshaw. 2018. 'The 'West Yorkshire Regional English Database': investigations into the generalizability of reference populations for forensic speaker comparison casework.' In *Proceedings of Interspeech 2018*, Hyderabad. 2748-52.
- Levi, A. M. 1998. 'Protecting innocent defendants, nailing the guilty: a modified sequential lineup', *Applied Cognitive Psychology*, 12: 265-75.
- McAleer, P., A. Todorov, & P. Belin. 2014. 'How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices.' *PLOS ONE*, 9(3), 9.
- McDougall, K., M. Duckworth, and T. Hudson. 2015. 'Individual and group variation in disfluency features: a cross-accent investigation.' In *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by The Scottish Consortium for ICPhS 2015, Glasgow, Paper number 0308.1-5. <http://www.icphs.info/pdfs/Papers/ICPHS08.pdf> Glasgow: University of Glasgow.

References 2



- McDougall, K., A. Paver, F. Nolan, N. Pautz, H.M.J. Smith and P. Harrison. 2021. 'Phonetic correlates of listeners' judgements of voice similarity within and across accents.' Paper presented at the International Association for Forensic Phonetics and Acoustics Annual Conference (online), Marburg, 22-25 August 2021.
- Nolan, F., K. McDougall, G. de Jong, and T. Hudson. 2009. 'The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research', *International Journal of Speech, Language and the Law*, 16: 31-57.
- Pozzulo, J.D., and R.C.L. Lindsay. 1999. 'Elimination lineups: An improved identification procedure for child eyewitnesses.', *Journal of Applied Psychology*, 84: 167-76.
- Seale-Carlisle, T.M., and L. Mickes. 2016. 'US line-ups outperform UK line-ups.', *Royal Society Open Science*, 3: 160-300.
- Smith, H.M.J., A. K. Dunn, T. Baguley, P. C. Stacey. 2016. 'Concordant Cues in Faces and Voices: Testing the Backup Signal Hypothesis', *Evolutionary Psychology*, 14: 1474704916630317.
- Smith, H.M.J., K. Bird, J. Roeser, J. Robson, N. Braber, D. Wright, and P.C. Stacey. 2020. 'Voice parade procedures: optimising witness performance', *Memory*, 28: 2-17.
- Stevenage, S.V., A. Howland, and A. Tippelt. 2011. 'Interference in eyewitness and earwitness recognition', *Applied Cognitive Psychology*, 25: 112-18.
- Stevenage, Sarah V., G. Clarke, and A. McNeill. 2012. 'The "other-accent" effect in voice recognition', *Journal of Cognitive Psychology*, 24: 647-53.
- Zimmermann, J.F., M. Moscovitch, and C. Alain. 2016. 'Attending to auditory memory', *Brain Research*, 1640: 208-21.