H32 The Reproducibility of Results From Facial Approximation Accuracy Tests That Use Face-Arrays

Carl N. Stephan, PhD, Joint POW/MIA Accounting Command Central Identification Laboratory, 310 Worchester Avenue, Hickam AFB, HI 96853-5530; Jody Cicolini, BSc, The University of Queensland, Brisbane, 4072 AUSTRALIA*

After attending this presentation, attendees will gain insights into the differences that can be expected when accuracy tests of facial approximations are conducted using different groups of assessors but the same face-array test.

This presentation will impact the forensic community by demonstrating the reliability of the most favored method for assessing facial approximation performance.

The use of facial approximation methods in the forensic arena makes knowledge of these methods accuracy and reliability pertinent. In the recent scientific literature it is now common for the face-array test to be advocated as the most favored form of performance assessment. If, however, the results from face-arrays are to be confidently interpreted then their reproducibility must be known. So far, repeatability studies using the same facial approximations but different assessor groups have not been performed. Clearly, more reliable results are expected with larger samples, but what numbers of assessors are needed and how reliable the results are for any sized assessor sample is currently unknown.

This study investigates the repeatability of face-array results for three previously published facial approximations by employing three new groups of assessors in addition to the original study's sample. In each case, face-arrays consisted of ten faces (including the target individual) with retest assessor sample sizes approximating 40, 75, and 115 individuals. The results show that fluctuations in the recognition rates up to 20% were not uncommon for single faces between samples, even when large assessor groups were employed (75-115 assessors). These data thereby suggest that recognition rates for single faces should only be considered to be approximate and that heavy weight should not be assigned to isolated recognition rates for single faces. If marked peaks are present in the recognition profile of the array, reliable indications to a facial approximation's accuracy may be gleaned from even small sampled studies, however, if such peaks are absent repeated tests using different assessor groups or a single large assessor samples (*at least* >75 individuals) are advisable. Since it appears that fluctuations in recognition rates between groups are not dramatically reduced with 75 to 114 assessors, it seems that much larger sample sizes might be needed to generate more reproducible results (but how large these samples must be is currently unknown). Future research must address this issue. Also, since recognition rates generated from face-arrays are expected to be heavily influenced by which foil faces are selected for inclusion into the array (in addition to the quality of the facial approximation), future research should examine methods for measuring and optimizing photo-spread fairness.

While face-array recognition tests are time intensive to conduct, the results of this study demonstrate value for obtaining larger assessor samples to run repeat face-array tests with multiple (independent) assessor groups. Repeat tests might be superfluous if very large numbers of assessors are used in any one pass at a face-array, but the exact number of assessors required to generate such robust results is currently unknown and could be unexpectedly large.

Facial Approximation, Facial Reconstruction, Facial Reproduction

 ** Presenting Author*