

Morrison, Epps, Rose, Thiruvaran, Zhang (2010):

## **Measuring Reliability in Forensic Voice Comparison**

*JASA* 128, 2378

### **ABSTRACT**

Recently there has been a great deal of concern in forensic science about validity and reliability (accuracy and precision). The log-likelihood-ratio cost ( $C_{llr}$ ), developed for automatic speaker recognition, is increasingly applied as a standard measure of accuracy in forensic voice comparison, but so far there has been little work on developing a metric of precision within this field. Because voice data can have a large amount of intrinsic variation at the source, and likelihood ratios are typically calculated using a single suspect recording and a single offender recording, assessing the precision of a forensic-voice-comparison system is extremely important. This presentation discusses the importance of measuring precision and describes two procedures, one parametric and one non-parametric, for calculating 95% credible intervals for the likelihood ratios resulting from running tests of forensic-voice-comparison systems (in which some comparisons are known to be same-speaker comparisons and others are known to be different-speaker comparisons). Examples are drawn from both acoustic-phonetic and automatic forensic-voice-comparison systems.