B26 An Approach to Conducting Automated Speaker Recognition

Hirotaka Nakasone, PhD, Federal Bureau of Investigation, Engineering Research Facility, Bldg 27958A, Quantico, VA 22135; Kenneth Marr, MS*, FBI, ERF, Bldg 27958A, Quantico, VA 22135; William L. Hinton, BS*, FBI, ERF, Bldg 27958A, Quantico, VA 22135; Alysha Jeans, BS*, FBI, ERF, Bldg 27958A, Quantico, VA 22135; and Ryan Lewis, BS*, FBI, ERF, Bldg 27958A, Quantico, VA 22135*

The goal of this presentation is to instruct participants regarding training and testing preparations prior to performing automated speaker recognition forensic examinations.

This presentation will impact the forensic science community by giving participants knowledge and insight into the training, testing, and procedures used at the FBI audio laboratory involving Automated Speaker Recognition (ASR) forensic examinations.

The automated speaker recognition technology has made significant progress in performance and accuracy in the past decade due to the rigorous and worldwide research effort. Although the performance and accuracy continues to improve, there also remain many unresolved challenges. An approach will be described to resolve such challenges in performing forensic voice comparison examinations by the use of an ASR system currently used at the Federal Bureau of Investigation for investigative use. The approach is described as an automated speaker recognition system with a human-in-the-loop which has channel-independent and language-independent capabilities. The ASR system is built on MIT Lincoln Laboratory's state-of-the-art algorithms i-Vector, Joint Factor Analysis, and Inner Product Discriminant Functions (IPDF) operated by a trained examiner. Some recent improvements have been made in the operating procedures, log Likelihood Ratio (LLR), score interpretations, test and validation methods for LLR score calibrations, examiner training curriculum, and an innovative automated human examiner training module. The final decision-making process uses five levels: "match," "probable match," "inconclusive," "probable no-match," and "no-match" based on the fused results of the ASR system and the human examiner.

The ASR examiner training consists of classroom sessions taught in-house and lectures provided by leading academic organizations, and self-paced peer-review sessions to establish whether trainees have mastered the skills and knowledge required to conduct ASR examinations. The entire training process was monitored for each trainee by a designated mentor and is documented using an ASR training curriculum. A six-month-long classroom instruction was conducted for examiner-trainees by a senior voice comparison examiner, consisting of basic audio signal processing, basic science of speech production, speech perception, phonetics, and statistics and probability theories. A specialized self-paced training program called the Forensic Training Module (FTM) is administered to each examiner-trainee to perform and successfully pass a set of more than 100 blind voice comparison tests under cross-language (the training speaker speaks language 1, while the test speaker speaks language 2), cross-channel (the training speaker speaks via telephone, while the test speaker speaks via microphone) by listening only. All voice samples in the databases used for the FTM training are truth marked, thereby providing the trainees immediate feedback for optimum learning purposes. It was determined that through the ASR development process, there are specific acoustic conditions which are encountered when conducting forensic voice comparison examinations. These conditions and the characteristics of each will be discussed.

The testing and validation of the ASR system was a culmination of years of development of automated tools, including Air Force Research Laboratory's RAPT-R system and Massachusetts Institute of Technology Lincoln Laboratory's Vocalinc system. A Test and Validation Plan was prepared and approved for the voice comparison analysis of thousands of known voices. The ground truth of the databases allowed LLR-score thresholds to be calibrated for five levels of decisions: match, probable match, inconclusive, probable no match, and no match for each of 14 separate acoustic recording conditions. Consequently, examiners use the ASR algorithms to calculate automated scores for questioned voice recordings. Research and testing to improve the ASR process continues, addressing other factors impacting the ASR performance including speech duration, signal-to-noise ratio, and the degree of reverberation in the environment.

The current policy and limitations of the use of ASR as a courtroom tool as well as some *Daubert* criteria challenges will be presented. Finally, the progress status of a new scientific working group for voice (tentatively called SWG-V) will be discussed; formulation of this group is under way and due to the joint efforts of the National Institute of Standards and Technology and the Federal Bureau of Investigation collaborating with scientific representatives from local, state, federal laboratories, academia, industry, and legal community.

ASR, Speaker ID, Voice Comparison

 ** Presenting Author*