

## Database selection for forensic voice comparison

*Geoffrey Stewart Morrison, Felipe Ochoa, Tharmarajah Thiruvanan*

Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications,  
University of New South Wales

geoff-morrison@forensic-voice-comparison.net

### Abstract

Defining the relevant population to sample is an important issue in data-based implementation of the likelihood-ratio framework for forensic voice comparison. We present a logical argument that because an investigator or prosecutor only submits suspect and offender recordings for forensic analysis if they sound sufficiently similar to each other, the appropriate defense hypothesis for the forensic scientist to adopt will usually be that the suspect is not the speaker on the offender recording but is a member of a population of speakers who sound sufficiently similar that an investigator or prosecutor would submit recordings of these speakers for forensic analysis. We propose a procedure for selecting background, development, and test databases using a panel of human listeners, and empirically test an automatic procedure inspired by the above. Although the automatic procedure is not entirely consistent with the logical arguments and human-listener procedure, it serves as a proof of concept for the importance of database selection. A forensic-voice-comparison system using the automatic database-selection procedure outperformed systems with random database selection.

---

Different portions of this paper have benefitted from two principal sources of financial support:

Sections 2–4: The Australian Research Council, Australian Federal Police, New South Wales Police, Queensland Police, National Institute of Forensic Science, Australasian Speech Science and Technology Association, and the Guardia Civil through Linkage Project LP100200142. Unless otherwise explicitly attributed, the opinions expressed are those of the authors and do not necessarily represent the policies or opinions of any of the above mentioned organizations.

Sections 5–7: The Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

Thanks to Alex Biedemann, Jason Pelecanos, anonymous reviewers and others for comments on earlier drafts of parts of this paper. The opinions expressed in this paper are those of the authors, and any shortcomings in the paper are the responsibility of the authors.

Thanks to the organizers of Odyssey 2012, COLIPS, and Temasek Laboratories@NTU for the invitation to submit the written version of this paper for publication in the Odyssey 2012 proceedings and to present the oral version in Singapore in conjunction with Odyssey 2012, including financial support for the latter from Temasek Laboratories@NTU.

### 1. Introduction

Difficulty in defining the appropriate population to specify in the defense hypothesis has been cited as a reason for not adopting data-based implementations of the likelihood-ratio framework for forensic voice comparison (French & Harrison [1]; French et al. [2]; but see responses in Rose & Morrison [3]; Morrison [4], Morrison [5] §9.400). In the current paper, we present a logical argument as to the appropriate population for the forensic scientist to sample for background, development, and test databases, and propose a human-listener procedure for selecting recordings to include in the sample. This is followed by a discussion of how the procedure would have been applied in three casework examples. We also discuss some potential objections to our proposed procedure. We leave empirical testing of the human-listener procedure for future research, but in the mean time describe and empirically test an automatic procedure inspired by the logical arguments and human-listener procedure. Although the automatic procedure is not fully consistent with the logical arguments, the results of tests of this procedure indicate that database selection does lead to better system performance.

### 2. Logical arguments

#### 2.1. A likelihood ratio is the answer to a specific question and this question specifies the relevant population

The aim of forensic voice comparison is to produce a likelihood ratio which is an expression of the strength of the evidence with respect to two competing hypotheses (Champod & Meuwly [6]; Rose [7]; Morrison [5]). The first hypothesis, the prosecution hypothesis, is usually that a voice of questioned identity on one audio recording (the questioned-speaker /offender recording) belongs to the same speaker as the voice on one or more other audio recordings for which the identity of the speaker is not disputed (the known-speaker / suspect recording). The alternative hypothesis, the defense hypothesis, is usually that the questioned voice does not belong to the suspect, but rather belongs to some other speaker. An appropriate defense hypothesis will, however, always be more specific than “some other speaker”, and the details of the defense hypothesis are part of the definition of the question which is answered by a likelihood ratio.

A likelihood ratio cannot be interpreted without