



INTERPOL survey of the use of speaker identification by law enforcement agencies



Geoffrey Stewart Morrison¹, Farhan Hyder Sahito^{2,*}, Gaëlle Jardine, Djordje Djokic, Sophie Clavet, Sabine Berghs, Caroline Goemans Dorny

Office of Legal Affairs, INTERPOL General Secretariat, 200 Quai Charles de Gaulle, 69006 Lyon, France

ARTICLE INFO

Article history:

Received 4 November 2015
Received in revised form 10 March 2016
Accepted 21 March 2016
Available online 1 April 2016

Keywords:

Survey
Speaker identification
Forensic voice comparison
Law enforcement
INTERPOL

ABSTRACT

A survey was conducted of the use of speaker identification by law enforcement agencies around the world. A questionnaire was circulated to law enforcement agencies in the 190 member countries of INTERPOL. 91 responses were received from 69 countries. 44 respondents reported that they had speaker identification capabilities in house or via external laboratories. Half of these came from Europe. 28 respondents reported that they had databases of audio recordings of speakers. The clearest pattern in the responses was that of diversity. A variety of different approaches to speaker identification were used: The *human-supervised-automatic approach* was the most popular in North America, the *auditory-acoustic-phonetic approach* was the most popular in Europe, and the *spectrographic/auditory-spectrographic approach* was the most popular in Africa, Asia, the Middle East, and South and Central America. Globally, and in Europe, the most popular framework for reporting conclusions was *identification/exclusion/inconclusive*. In Europe, the second most popular framework was the use of *verbal likelihood ratio scales*.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The present paper reports the results of a survey conducted as to the status quo in law enforcement agencies' use of speaker identification. This survey was conducted as an initial step in a larger project aimed at developing investigative and police intelligence applications of speaker identification technology. Knowledge of the status quo may help in understanding the potential needs of law enforcement agencies in this area.

2. Methodology

A questionnaire was circulated to law enforcement agencies in INTERPOL's 190 member countries, and the responses analysed.

The questionnaire underwent several rounds of drafting and revision. It was revised based on feedback from members of the larger project's research consortium, and from seven forensic scientists with working knowledge of forensic voice comparison and/or forensic audio analysis. Each of the latter worked for a law enforcement agency, a judicial agency, or a private laboratory.

Feedback was also provided by an academic researcher who specialises in automatic speech recognition, and by a staff member of a standards organisation.

The final version of the questionnaire was written in English and translated into INTERPOL's other three working languages, French, Spanish, and Arabic. A copy of the English version of the questionnaire appears in the appendix to the present paper. The questionnaire was sent to INTERPOL's National Central Bureaus in each of the 190 member countries. The National Central Bureaus were asked to send the questionnaire to the agency or agencies within their country which they deemed best able to respond to the questionnaire. They were asked to consider sending the questionnaire to departments responsible for "cybercrime; counterterrorism units; forensic laboratories and other experts; [and] operational units". Of responses received from law enforcement agencies which reported having speaker identification capabilities, most of those responses were prepared by personnel associated with forensic laboratories.

The research team was able to immediately analyse responses written in English, French, and Spanish. Responses written in Arabic were translated into English by professional translators prior to analysis. The research team included members with expertise in evaluation of forensic evidence, in forensic voice comparison, in law enforcement, and in law. Their expertise was used in interpreting the answers to the questions asked by the questionnaires. For example, some questions had multiple

* Corresponding author.

E-mail address: f.sahito@interpol.int (F.H. Sahito).

¹ Corresponding author during submission and review process

² Author to whom correspondence regarding the published paper should be addressed

preformatted answers plus an “other” option with a space for a comment. Based on their expertise, team members could often determine that the answer provided in the comment was actually within the scope of one of the preformatted answers and simplified analysis by selecting the relevant preformatted answer.

In reporting results of the survey, we focus on what we consider to be noteworthy observations rather than providing a detailed description of the answers to all questions asked.

3. Results and discussion

3.1. Origin of responses

A total of **91 responses were received from 69 countries**. Two or more responses were returned from some countries. Half the responses came from Europe.

38 respondents said that their law enforcement agency had a laboratory which analysed voice recordings (one said they also used an external laboratory), and another 6 said they used external laboratories. Thus **44 respondents’ agencies had the capability to analyse voice recordings, either internally or externally**. The regional distribution of these respondents is shown in Fig. 1.

Not all questions were necessarily answered in each response. The total number of answers per question therefore varied from question to question. Responses were provided on the condition that particular law enforcement agencies and countries would remain anonymous. Results below are therefore pooled. They are pooled either across all responses, or, where there are noteworthy differences between regions, they are pooled within each region.

3.2. Legal context

Among respondents whose agencies have speaker identification capabilities, substantial differences were reported with respect to the legal frameworks within which they operate. In most cases there is no legislation specifically addressing the use of speaker identification. Requirements which would impose limitations on the use of speaker identification are scattered across different laws and regulations, including those relating to privacy and data protection, criminal procedures, and judicial authorisation.

Respondents reported using speaker identification technology to address a broad range of categories of crime, with 21 to 30 positive responses for each of the following categories: cyber, corporate/white-collar, hate, identity, organized, other.

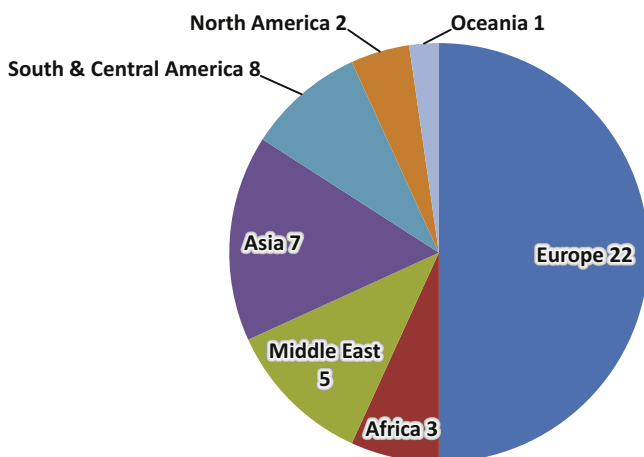


Fig. 1. Regional origins of respondents who reported having speaker identification capabilities.

3.3. Speaker identification technology

Some respondents reported that their agencies used forensic voice comparison systems or automatic speaker recognition systems:

- 12 reported using BATVOX (produced by Agnitio, a company based in Spain);
- 5 reported using IKAR Lab (produced by Speech Technology Center, a company based in Russia);
- 3 reported using Loquendo Voice Investigation System (LVIS, produced by Loquendo, a company based in Italy, it is now part of Nuance);
- 2 reported using Vocalise (produced by Oxford Wave Research, a company based in the United Kingdom);
- 1 reported using Phonexia Speaker Identification (produced by Phonexia, a company based in the Czech Republic);
- 1 reported using ALIZE (open source software produced by researchers at the University of Avignon in France);
- 2 reported using forensic voice comparison systems which had been developed in house (a different system used by each respondent).

Respondents also reported using general speech and audio analysis tools, including hardware and software produced by KayPENTAX (formerly Kay Elemetrics), Adobe Audition software (or an earlier version, Cool Edit Pro produced by Syntrillium), and Praat software (freeware produced by researchers at the University of Amsterdam in the Netherlands).

3.4. Databases of voice recordings

Of the 44 respondents who reported that their agency had the capability to analyse voice recordings, 26 indicated that they had a database. Thus less than 60% of the respondents who had the capability to analyse voice recordings also had a database of voice recordings. 2 additional respondents also said they had databases even though they did not have the capability to analyse voice recordings.

Of the **28 respondents total who reported having databases**, most (20) reported that those databases contained *suspect* recordings, and less than half (10 and 13 respectively) reported that they contained recordings of *convicted persons* and *population samples* (persons unconnected to any particular crime). Only 4 reported having databases which included *missing persons*, and 1 mentioned recordings of *terrorists*. Offenders of unknown identity was not a category included in the preformatted answers and not explicitly added by respondents, although it may have been included in the *crime scene recordings* reported in the next paragraph.

The reported sources of speaker recordings were relatively evenly spread across *crime scene recordings* (exact details of recording device not specified), *landline telephone* intercepts, *mobile telephone* intercepts, *interviews* with suspects, and *open source* (e.g., from the internet), each of these being reported by 11 to 14 respondents. Voice over internet (VoIP) recordings were reported by 6 respondents. 5 respondents reported having recordings specifically collected as population samples or for research. The latter could potentially have included high-quality audio recordings.

Less than half the respondents (12) reported that their databases had multiple recordings of the same speaker. Of those that had multiple recordings per speaker the modal number was 3, with one outlier reporting 25. Recordings were reported to have durations ranging from 30 s to 40 min, with a mode of 2 min. The number of recording channels (landline, mobile telephone, etc.) included in each database ranged from 1 to 5. The number of

Download English Version:

<https://daneshyari.com/en/article/95142>

Download Persian Version:

<https://daneshyari.com/article/95142>

[Daneshyari.com](https://daneshyari.com)