Letter to the Editor

Refining the relevant population in forensic voice comparison – A response to Hicks *et alii* (2015) The importance of distinguishing information from evidence/observations when formulating propositions

Geoffrey Stewart Morrison [a,b,*], Ewald Enzinger [a], Cuiling Zhang [c,d]

[a] *Morrison & Enzinger, Independent Forensic Consultants, Vancouver, British Columbia, Canada & Corvallis, OR, USA*
[b] *Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada*
[c] *School of Criminal Investigation, Southwest University of Political Science and Law, Chongqing, China*
[d] *Chongqing Institutes of Higher Education Forensic Technology Research Centre, Chongqing, China*

ABSTRACT

Hicks *et alii* [Sci. Just. 55 (2015) 520–525. http://dx.doi.org/10.1016/j.scijus.2015.06.008] propose that forensic speech scientists not use the accent of the speaker of questioned identity to refine the relevant population. This proposal is based on a lack of understanding of the realities of forensic voice comparison. If it were implemented, it would make data-based forensic voice comparison analysis within the likelihood ratio framework virtually impossible. We argue that it would also lead forensic speech scientists to present invalid unreliable strength of evidence statements, and not allow them to conduct the tests that would make them aware of this problem.

© 2016 The Chartered Society of Forensic Sciences. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

A forensic likelihood ratio can be expressed as $LR = p(E|H_p,I)/p(E|H_d,I)$, in which LR is the likelihood ratio, E is the evidence, H_p is the prosecution hypothesis, H_d is the defence hypothesis, I is background information about the circumstances of the case, and $p(E|H,I)$ is the probability of the evidence given a hypothesis and the background information. Hicks *et alii* [1] discuss the importance of not confusing the evidence with background information. In $p(E|H,I)$, no part of E, which is to the left of the conditioning bar, should be included in H or I, which are to the right of the conditioning bar. Hicks *et alii* provide a number of useful didactic examples illustrating problems resulting from making such an error.

In general, we think that Hicks *et aliorum* didactic goals are laudable, but in their *Example 5 (speech analysis)*, Hicks *et alii* are writing outside their own areas of expertise. None of them are experts in forensic voice comparison, and they do not even cite any literature specifically related to this field. Speech is a complex phenomenon, and Hicks *et alii* take an overly simplistic view which treats speech evidence as a nondecomposable whole. Lack of understanding of the realities of forensic

voice comparison has led them to make a proposal which, if left unchallenged, could be detrimental to progress in this branch of forensic science.[1] If followed, the proposal would unnecessarily make data-based forensic voice comparison analysis within the likelihood ratio framework almost impossible, and below we will argue that it would lead forensic speech scientists to present invalid unreliable strength of evidence statements without realising that they are doing so. We will empirically demonstrate the problem with respect to validity. There is also the danger that lawyers could reference Hicks *et alii* [1] in order to obfuscate a forensic speech scientist's logically correct presentation of strength of evidence. Forensic voice comparison has been one of the most progressive branches of forensic science in terms of adopting the logically correct framework for the evaluation of evidence, calculating likelihood ratios on the basis of relevant data, quantitative measurements, and statistical models, and testing validity and reliability. This branch of forensic science, however, is still in the midst of a paradigm shift, and those who work in the new paradigm are a minority who

---

[1] We do not mean to imply that there is something special about forensic voice comparison which makes it different from all other branches of forensic science. It may be that arguments presented here are also applicable in some other branches of forensic science, but we refrain from making arguments specific to branches of forensic science which are outside our expertise.

* Corresponding author.
E-mail address: geoff-morrison@forensic-evaluation.net. (G.S. Morrison).

continue to face concerted opposition. In this context, Hicks *et aliorum* proposal is not helpful, to say the least. As proponents of the new paradigm, and practising forensic speech scientists who have thought long about these matters and their implications for practice, we feel obliged to respond. Below we describe in detail why we believe that Hicks *et aliorum* proposal should not be followed.

2. Hicks *et aliorum* example

Hicks *et aliorum* example is as follows: In the perception of a police officer, the recording of the speaker of questioned identity exhibits a particular regional accent and the speaker of known identity has the same regional accent. The prosecution hypothesis is that the questioned speaker is the known speaker. If the police officer's conclusions regarding accent are not contested by the defence, then Hicks *et alii* say that it "might be appropriate" to adopt the defence hypothesis that the questioned "speaker is an unknown person from the same region" (we would phrase this as "a person selected at random from the relevant population, where the relevant population is people who speak with the particular regional accent"). If, however, the police officer's conclusions are not presented to the court, Hicks *et alii* state that a more appropriate hypothesis would be that the questioned "speaker is an unknown person. To elicit who is this unknown person [*sic*], one will have to take into consideration the relevant population in this case, depending on the case circumstances. These may dictate a given region, but the voice itself should not drive this decision." We assume that Hicks *et alii* intended "voice of the questioned speaker", and that we all agree that information about the known speaker should play no part in defining the relevant population specified by the defence hypothesis (Robertson & Vignaux [2] pp. 43–44).

Hicks *et alii* equate with *E* all information which could be derived by the forensic speech scientist from the recorded speech of the speaker of questioned identity. Thus, circumstances of the case known for other reasons may be used to refine the relevant population (*I* can be used to refine H_d), but any information that the forensic speech scientist derives from the speech of the questioned speaker must not be used to refine the relevant population (no part of *E* can be used to refine H_d).

Although the example was based on a regional accent, note that other classes or pseudo-classes could be relevant, for example, language spoken, male versus female, or deep voice versus high-pitched voice. In reality these are not always clearly differentiable classes, but treating them as such simplifies discussion.

3. Scenarios for the submission of speech recordings for forensic analysis

There are different scenarios for the submission of known- and questioned-speaker recordings for forensic analysis.

One scenario (as in the Hicks *et aliorum* example, and also discussed in detail in Morrison *et alii* [3]) is that the decision to submit the known- and questioned-speaker recordings for analysis comes after someone such as a police officer or a prosecutor (not an expert in forensic speech science) has listened to them and decides they are sufficiently similar sounding that it is worth submitting them for forensic analysis. If the listener does not think they are sufficiently similar sounding, they will not generate the same-speaker hypothesis, and they will not submit the recordings for analysis. This is very different from other types of forensic evidence such as DNA, where the person submitting specimens for analysis has no access to information about their properties. Note that in this scenario the decision to submit is based on the listener's perception, and, unbeknownst to the listener, the known and questioned speakers could speak with two different accents that are unfamiliar to the listener (Labov & Harris [4]).

A second scenario is a variant of the first. A single individual could speak two different languages or could speak with two different accents,

or could disguise their voice, or differences in speaking style and recording conditions could make the two recordings sound quite different. A non-expert listener is aware of such differences but still believes that the two recordings were produced by the same speaker – they sound sufficiently similar to the listener despite these differences.

A third scenario is that the decision to submit the recordings for analysis is motivated by other considerations, and is not filtered by a non-expert listener's perception. In this scenario, the known- and question-speaker recordings could sound quite different from one another. This is actually quite common.

In our practice as forensic speech scientists, we begin with a preliminary analysis, which includes an assessment of the recording conditions and speaking styles on the known- and questioned-speaker recordings. If we notice a mismatch in language or accent, we contact the client to ascertain if a cross-language or cross-dialect hypothesis is reasonable. If it is not, and the client is the prosecution, then they may decide to pursue other suspects or other theories of the case, and are unlikely to want us to continue with a full analysis. If the hypothesis is reasonable, and we proceed with the analysis, then we will have to train and test our system under these mismatched conditions (see Enzinger *et alii* [5] for an example of training and testing a forensic voice comparison system under mismatched conditions).

4. Refining the relevant population using questioned-speaker accent avoids double counting

It is also during the preliminary analysis that we refine the prosecution and defence hypotheses that we will adopt, including the definition of the relevant population. Before proceeding with the full analysis we confer with the client, and other parties as may be appropriate. We adopt hypotheses which pose a particular question. We do our best to make sure that that question is one which will potentially have an answer which will be of use to the trier of fact in determining a relevant issue in the case (Morrison & Stoel [6]).

Whether the submission of the recordings was filtered by a non-expert listener or not, it will usually be obvious to a forensic speech scientist whether the questioned speaker is male or female, what language they are speaking, and broadly what accent they are speaking. These properties will usually also be perceptually salient to all parties (prosecution, defence, judge, and trier of fact), either immediately or after they have been pointed out by the forensic speech scientist. If this is the case, speaker sex, and language and accent spoken can be taken as background information *I*, not as evidence *E*. They can therefore be used to refine the relevant population specified in the defence hypothesis (Rose [7] pp. 64–65).

Conducting an analysis based on a relevant population refined in this way will actually help avoid double counting. There are serious problems with speaker recognition by laypersons (see reviews in Solan & Tiersma [8], Morrison [9] §99.910–99.1110, Edmond *et alii* [10], Sherrin [11]), but the courts invariably permit the trier of fact to listen to recordings (Edmond *et alii* [12]). So in addition to being presented with the results of the forensic speech scientist's analysis of the strength of the speech evidence, the trier of fact will also be exposed to and influenced by perceptually salient properties such as speaker sex, and language and accent spoken.

5. Not using questioned-speaker accent to refine the relevant population leads to invalid unreliable strength of evidence conclusions

Most Australians can easily distinguish New Zealand and Australian accents. Thus in an Australian court the fact that the questioned speaker has a New Zealand accent can usually be taken as part of *I*, and the relevant population can be specified as New Zealand English speakers. The trier of fact will hear for themselves that the questioned speaker has a New Zealand accent and be interested in whether the questioned