Critique by Dr Geoffrey Stewart Morrison of a forensic voice comparison report submitted by Mr Edward J Primeau in relation to a section of audio recording which is alleged to be a recording of the voice of Dr Marlo Raynolds

This document was prepared at the behest of:

Ms Joan Bryden, Parliamentary Reporter, The Canadian Press

Prepared by:

Geoffrey Stewart Morrison BSc MTS MA PhD

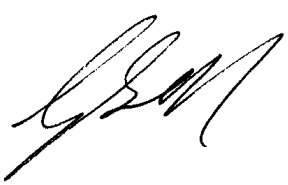
http://geoff-morrison.net/

http://forensic-evaluation.net/

http://forensic-voice-comparison.net/

This version completed:

30 November 2014



Digitally signed by Dr Geoffrey Stewart Morrison
Date: 2014.11.30 20:22:09 -08'00'

Table of Contents

Note: Page numbering excludes appendices.

1 Background

1.1 On 27 November 2014 I was alerted to a newspaper article regarding a forensic voice comparison case (Bryden, 2014).

1.2 The case involves an audio recording. In the latter part of the recording (after 1 minute and 58 seconds into the recording, henceforth 1:58) the Conservative Party of Canada alleges that the speaker is Dr Marlo Raynolds, a Liberal Party of Canada candidate for election to the House of Commons. Dr Raynolds and another individual, Mr Tam McTavish, claim that the speaker is Mr McTavish. Both sides agree that Dr Raynolds is speaking at the beginning of the recording, but disagree as to who is speaking after 1:58.

1.3 The Conservative Party commissioned Mr Edward J Primeau to perform a forensic analysis of the audio recording. Mr Primeau represents himself as "an audio and video forensic expert" (Primeau report). Mr Primeau concluded that he is "100% sure that the male at the beginning of the digital audio recording, and the male speaking at the one minute and fifty eight second mark, is the same person."

1.4 Based on the information provided in the newspaper article, I was concerned about the quality of the forensic analysis conducted in that case. I contacted the journalist who had written the newspaper article, Ms Joan Bryden. She offered to provide me with a copy of the forensic report and the audio recording, and I agreed to write a critique of the forensic report. The current document is that critique.

2 Qualifications and Experience

2.1 Highlights of my qualifications and experience are provided in the paragraphs below, and are detailed in my Curriculum Vitae in Appendix A.

2.2 I received my PhD from the Department of Linguistics, University of Alberta in 2006. My dissertation focussed on statistical modelling of speech data. I began work on

forensic voice comparison in 2007 when I was appointed Research Associate on a project directed by Dr Philip Rose at the School of Language Studies, Australian National University. From 2010 to 2013 I was the Director of the Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales. Over approximately the same time period I was also an Invited Lecturer in the Judicial Phonetics Specialisation, Master in Phonetics and Phonology Programme, Spanish National Research Council / Menéndez Pelayo International University, and was Chair of the Forensic Acoustics Subcommittee, Acoustical Society of America. I am currently a Visiting Researcher at the School of Electrical Engineering & Telecommunications, University of New South Wales, an Adjunct Associate Professor, Department of Linguistics, University of Alberta, and a Subject Editor for the refereed journal Speech Communication.

2.3 I have been the lead researcher on research projects valued at over $900 000, including an Australian Research Council Linkage Project in collaboration with the Australian Federal Police, New South Wales Police, Queensland Police, National Institute of Forensic Science, Australasian Speech Science & Technology Association, and the Guardia Civil.

2.4 I have published a number of articles on forensic voice comparison and on the evaluation of forensic evidence in refereed journals in acoustics, speech processing, forensic science, and law. I am also the author of an introduction to forensic voice comparison in the Expert Evidence series (Morrison, 2010). I have been an invited speaker at several academic conferences, and have also given tutorial workshops on evaluation of forensic evidence at a number of academic conferences and at operational police laboratories.

2.5 At the 21st International Congress on Acoustics in Montréal in 2013, I was organiser of a special session entitled “Distinguishing between science and pseudoscience in forensic acoustics” <http://montreal2013.forensic-acoustics.net/>. A refereed journal article version of my presentation at that event is published as Morrison (2014).

2.6 I have previously provided written expert reports in relation to 14 legal cases (7 at the behest of the prosecution/plaintiff and 7 at the behest of the defence/respondent), contributed to 2 amicus briefs, and have given expert testimony in court on 4 occasions.

3 Expert Witness Code of Conduct

3.1 I have read a copy of the Federal Court of Canada Code of Conduct for Expert Witnesses [Schedule (Rule 52.2) of Federal Court Rules SOR/98-106, current to 2014-11-11 and last amended 2013-08-08], and agree to be bound by it or a parallel provincial or territorial code of conduct as may be applicable. A copy of the Federal Court of Canada Code of Conduct is provided as Appendix B.

3.2 I am a member of the International Association for Forensic Phonetics and Acoustics and agree to be bound by its code of practice, a copy of which is provided as Appendix C.

4 Quality Control

4.1 An earlier version of this report has been proofread by my colleague, Mr Ewald Enzinger, and I have taken account of their feedback in producing the final version. Mr Enzinger is a PhD candidate, School of Electrical Engineering & Telecommunications, University of New South Wales.

5 Disclaimers

5.1 Any opinions I express are my own and do not necessarily reflect the policies of any organisations with which I am affiliated, or the opinions of any individuals with whom I am associated.

5.2 The current document is a critique of Mr Primeau's report. At this stage I have not performed a forensic voice comparison myself and will not speculate on the exact strength of evidence I would obtain if I were to perform a forensic voice comparison in this case.

6 Materials - Primeau Report

6.1 I received from Ms Bryden via e-mail on 27 November 2014 a Microsoft Word format version of the forensic voice comparison report written by Mr Primeau. The report is three pages long, formatted as a letter, and addressed to Mr Michael Oberman.

6.2 I assume that the document which I received accurately reflects the report that Mr Primeau submitted. If this assumption is incorrect then what I write in the current document may need to be revised accordingly.

6.3 The report is publically available from the following URL:

http://www.conservative.ca/release_files/AnalysisReport.pdf

6.4 The only substantial difference of which I am aware between the text of the version send to me by Ms Bryden and the publically available version is the redaction of Mr Oberman's address in the latter.

6.5 The publically available version of Mr Primeau's report is appended to the current critique as Appendix D.

7 Materials - Audio Recording

7.1 I received from Ms Bryden via e-mail on 27 November 2014 a link to an audio recording: https://www.youtube.com/watch?v=0c8S41naH1k

7.2 I understand that the audio recording at this URL is a copy or a version of the audio

recording in question, referred to as "Audio Recording.MP4" in Mr Primeau's report. If this is incorrect then what I write in the current document may need to be revised accordingly.

7.3 Text on the website at the URL above includes the following: "Marlo Raynolds in Canmore, AB. Carry Jovi. Published on 22 Nov 2014. Handing people money for arguably nothing. Look at what happened in Alberta: People didn't go and invest in the economy, or invest it in businesses, or invest it in something. They bought TVs, they bought cars, they bought, you know, all these different things. They purchased items that really didn't do anything to benefit the economy."

7.4 If I were to perform a forensic voice comparison in this case I would first need to be provided with a direct copy of the original file, either in the original format or converted to an uncompressed format along with details of the original format and the software and procedures used for conversion.

7.5 Bryden (2014) states that "The audio, recorded during a Nov. 13 public meeting in Canmore, Alta., is a little over two minutes long and of varying quality. It is clear that at least two men and one woman are involved in the conversation about income splitting. The relevant remarks are toward the end, by which time the audio is muffled and the voices partially obscured by background noise. Lilley has told Raynolds the poor quality of the audio was due to the female Conservative supporter 'zipping up her coat when she thought you were getting nervous that she might be recording you.'" I will assume that the facts reported in this quotation are not disputed by either party. If this assumption is incorrect then what I say in the current document may need to be revised accordingly.

7.6 The following text in this subsection is intended only as a subjective description and does not constitute a professional opinion should any of what I write in this subsection be in dispute. Upon listening to the recording I note a type of background noise on the recording known as multi-speaker babble, i.e., multiple speakers talking in the background. The quality of the recording of the three interlocutors appears to be

relatively good at the beginning of the recording. At about 1:47 there is a noise which could be a zipper being closed, after which the quality of the audio deteriorates in a manner could be due to clothing covering the microphone. Between about 1:47 and 2:05 there is also a higher incidence of transient noises which could be due to clothing impacting or rubbing against the microphone. The quality of the recording of the interlocutors appears to be relatively bad in the latter part of the recording.

7.7 The dispute between the parties in this case is over the identity of the male speaker from 1:58 to the end of the recording.

8 How a Forensic Voice Comparison Should Be Conducted in this Case

8.1 In order to provide a contrast with Mr Primeau's report, the current section describes how I would conduct a forensic voice comparison in this case.

8.2 It is agreed among a large number of individuals with expertise in the field of evaluation of forensic evidence that the logically correct way to quantify the strength of forensic evidence is via a likelihood ratio (Robertson & Vigneaux, 1995; Rose, 2002; Aitken & Taroni, 2004; Balding, 2005; Association of Forensic Science Providers, 2009; Aitken et al, 2010; Evett, et al, 2010; Berger et al, 2011; Redmayne et al, 2011; Robertson et al, 2011; Morrison, 2012; Morrison et al, 2012). This is also the policy of several organisations including the Netherlands Forensic Institute and the European Network of Forensic Science Institutes. The likelihood-ratio framework is standard for the forensic evaluation of DNA (Foreman et al, 2003) and is gradually being adopted for the forensic evaluation of voice recordings (Morrison, 2009; Grigoras et al, 2013).

8.3 At the core of the likelihood-ratio framework is the understanding that it is not enough to know how similar an offender sample and a suspect sample are, one also has to know how typical the offender sample is. For example, let us say that a crime was committed and all eye witnesses agree that the offender had blond hair. A suspect is

arrested and the suspect also has blond hair. The similarity between the suspect and offender on this feature is high. Let us simplify by assuming that blond is clearly distinct from other hair colours, that no wigs were worn, that the eye witnesses are not mistaken, etc. The suspect has blond hair and the offender has blond hair. What is the probability of the offender having blond hair if the offender were the same person as the suspect? Given all simplifications, the answer you gave to this question was probably 100% (laypeople typically quantify probabilities using percentages which range from 0 to 100, whereas statisticians quantify probabilities as values between 0 and 1, the conversion is a trivial multiplication or division by 100). Now, does this mean that the suspect and the offender are the same person? You probably answered "no" to this question because you know that quite a lot of people have blond hair. The suspect is one person with blond hair so they could be the same person as the offender, but some other person with blond hair could also be the offender. Is the evidence about hair colour therefore useless? This might take you longer to think about, but I would answer that it is not useless, it has some value. A lot of people have blond hair, but a lot of other people don't. Given the simplifications, we can say that the offender must have been one of the people in the population who has blond hair, and we can rule out all of the people in the population who don't have blond hair. What is the probability of having blond hair? You shouldn't answer this question straight away, instead you should ask me a clarification question: Given what population? What I should really have asked was what is the probability of having blond hair in the relevant population? So, what is the relevant population? Well, this depends where the crime was committed. In most places in Canada, blond hair is neither really rare nor really common, but what if the crime had been committed in Stockholm? Blond hair is really common there, so the probability of having blond hair in that population is very high. What is the crime had been committed in Beijing? Blond hair is really rare there, so the probability of having blond hair in that population is very low. So it is important to decide what the relevant population is in this case. Let's decide that the relevant population is the one living in the region where the crime was committed. If the offender is not the suspect then they must have been somebody else living in the

region (we could make it an island and say noone had arrived or left since the crime was committed). There are about a million people living in this region and we can't look at all of them, but we can go out on the streets and note the hair colour of a few hundred people (this is a sample of the population). Of the people in our sample we find that about 20% have blond hair, and assuming that our sample is representative of the whole population we estimate that about 20% of the people in the population have blond hair. This is how typical blond hair is in the relevant population. The offender had blond hair, how probable is this evidence if the offender were someone from the relevant population? Answer: 20%. Now, we are ready to quantify the strength of the evidence. The evidence is that the offender has blond hair. How probable is the evidence if the offender were the suspect, versus how probable is the evidence if the offender were someone else from the relevant population? Or to put it another way: How similar is the offender with respect to the suspect, versus how typical is the offender with respect to the relevant population? Given the numbers we came up with earlier, 100% and 20%, the answer is that the evidence is 5 times more probable if the offender were the suspect than if the offender were someone else from the relevant population (100/20 = 5). This is our likelihood ratio, this is our quantitative estimate of the strength of the evidence.

8.4 In the present case, the evidence is the acoustic properties of the male voice on the recording after 1:58.

8.5 The first question to be answered is: What is the probability of the evidence if it had been spoken by Dr Raynolds.

8.6 The second question to be answered is: What is the probability of the evidence if it had been spoken by Mr McTavish.

8.7 The likelihood ratio which will quantify the strength of the evidence in this case is: The probability of the acoustic properties of the voice on the recording after 1:58 if it had been spoken by Dr Raynolds, versus the probability of the acoustic properties of the voice on the recording after 1:58 if it had been spoken by Mr McTavish.

8.8 This is a relatively simple case in that the relevant population consists of only one speaker, Mr McTavish, so we do not need to get voice samples from a large number of speakers to represent an even larger population. Further, if there is no dispute that Dr Raynolds is speaking during particular portions of the first part of the recording and that Mr McTavish is speaking during other particular portions of the first part of the recording, then we can extract acoustic properties from these portions of the recording and use them to build statistical models which will allow us to calculate the probability of the evidence given the two competing hypotheses, that Mr Reynolds spoke the evidence versus that Mr McTavish spoke the evidence.

8.9 There is a complication in that the quality of the audio recording of the evidence (from the last part of the recording) differs from the quality of the audio recording of the data (from the first part of the recording) that we would use to build the statistical models. In general, such differences in recording conditions tend to give us likelihood ratios which are closer to a value of 1, i.e., closer to the evidence being equally probable for both hypotheses. If there were a mismatch between the recording conditions for the sample of Dr Raynolds' speech and the sample of Mr McTavish's used to build the statistical models, then this would be expected to introduce a bias into the calculation of the likelihood ratio which may favour one or the other of the hypotheses. Since the conditions for recording Dr Raynolds and Mr McTavish are the same, those at the beginning of the recording, we are not concerned that there would be such a bias. Despite the mismatch between the recording conditions of the undisputed samples of Mr Reynold's and Mr McTavish's voices versus the recording conditions for the sample of disputed identity, the former were recording on the same microphone and recording device close in time, and are therefore expected to be more closely matched than if any other recordings of these speakers were used.

8.10 If I were to perform a forensic voice comparison on this material, I would provide a detailed description of both the acoustic and the statistical analyses performed. The aim would be to provide sufficient detail that another suitably qualified and equipped person could replicate what I did. The conditions of this case have some similarities

with a case I previously worked on. A research paper based on the conditions of that case (Enzinger & Morrison, 2014) provides details of acoustic and statistical analyses similar to those that would be applied in the current case.

8.11 Once I had extracted the acoustic information and built the statistical models, I would test the performance of the system I had built. Unless one tests the degree of validity and reliability of a system under conditions reflecting those of the case under investigation, then one does not know how well the system will perform under those conditions. One would not know whether the system is highly valid and reliable, or whether its performance is no better than chance. The lack of testing of validity and reliability in forensic science was a major concern in the 2009 National Research Council Report to the US Congress (National Research Council, 2009). Morrison (2011) describes procedures for testing the validity and reliability of forensic likelihood ratio systems. Morrison (2014) reviews calls over the last five and a half decades for the validity and reliability of forensic voice comparison to be tested under conditions reflecting those of the cases to which they are applied.

8.12 In the present case, portions of Dr Raynolds' data and portions of Mr McTavish's data could be removed from the data used to train the statistical models, the models could be trained without these data and then used to calculate likelihood ratios for the withheld data (this is a process known as cross validation). The value of the likelihood ratio would then be compared with knowledge as to whether the data being tested came from Dr Raynolds or Mr McTavish. All of these training and testing data, however, would come from the first part of the recording and none would have the same condition mismatch as when the actual disputed data from the last part of the recording is evaluated. A way to investigate the effect of the mismatch would be to record pairs of other speakers under conditions reflecting those of the current case and compare the results of testing on these pairs of speakers without the mismatch and with the mismatch. Such an approach was adopted in Enzinger & Morrison (2014).

8.13 The last thing I would do, after building and testing the statistical models, would be to evaluate the strength of the actual evidence in the case.

9 Mr Primeau's Report

9.1 Mr Primeau's report addresses two questions: Whether "the male speaking in the beginning of the recording matched the voice of the male speaking after the one minute and fifty eight second mark." And whether "the male in the YouTube video is the same male that is speaking in both portions of the digital audio recording." The "YouTube video" having been identified as a recording of Dr Raynolds. Given that both parties are in agreement that the speaker at the beginning of the first audio recording is Dr Raynolds, the second question is moot, and we need only be concerned with the first question.

9.2 Mr Primeau states:

> "The goal of a voice identification test is to compare the known and unknown voices using critical listening, electronic measurement and visual inspection of sound wave formation and color sonogram.
>
> This report will include descriptions of the similarities observed during critical listening, electronic measurement and visual inspection testing."
>
> "Using critical listening skills that I developed over the last 30 years I listened to the known and unknown voices repeatedly until I was familiar with all voice recordings.
>
> I then created a composition audio file placing speech samples back to back so that I could critically listen to the similarities as well as the differences."

9.3 The report does not include any examples of "electronic measurement".

9.4 The approach which Mr Primeau employed is known as the aural-spectrographic approach. Descriptions of this approach can be found in multiple publications (including: Tosi, 1979; National Research Council, 1979; Gruber & Poza, 1995; Morrison, 2010). The approach involves listening to the audio recording of the voice

of known identity and of the voice of questioned identity. It also involves looking at pictures which represent the acoustic properties of the signal. Mr Primeau reports looking at waveform representations (which he calls "sound wave formation"), which represents time on the *x* axis and instantaneous amplitude on the *y* axis, and spectrographic representations, which represent time on the *x* axis and frequency on the *y* axis and represent amplitude as degree of darkness on a monochrome scale or as different colours on a colour scale. Mr Primeau reports using colour spectrograms (which he calls "color sonogram"). On the basis of auditory and visual perception, the practitioner forms a subjective judgement as to whether the samples of known and questioned identity were spoken by the same person or by different people.

9.5 Almost since its inception in the 1960's the aural-spectrographic approach (and the visual only spectrographic approach) has been criticised for being overly subjective and susceptible to bias, and for its degree of validity and reliability not having been empirically demonstrated under conditions reflecting casework conditions (see review in Morrison, 2014).

9.6 No approach to evaluation of forensic evidence is entirely objective. There is always some degree of subjectivity. In the case of the aural-spectrographic approach the final decision as to the weight of the evidence is a subjective judgement. This makes it relatively highly susceptible to cognitive bias, i.e., unconscious influence from domain irrelevant information (Dror & Stoel, 2014). My preference (as argued in multiple publications including: Morrison, 2014; Morrison & Stoel, 2014) is to employ an approach based on relevant data, quantitative measurements, and statistical models. Subjective decisions are made in deciding on matters such as the relevant hypotheses to test, selecting a sample of data that is expected to reflect the relevant population, and deciding the specific acoustic measurements to make. Such decisions are, however, far removed from the final calculation of the strength of the evidence. In my approach the quantitative-measurement and statistical-model system is first built and tailored to the particular questions and conditions of the case under investigation. The system is then frozen – no subsequent changes are allowed. The validity and reliability

of the system is then empirically tested under conditions reflecting those of the case under investigation, and the results of such testing recorded. Finally, the last thing which is done is to evaluate the strength of the actual evidence in the case using the quantitative-measurement statistical-model system. The value reported for the strength of the evidence is the value output by the statistical model. The subjective decisions are far removed from this final calculation of the strength of the evidence, and the value derived is less susceptible to influence from cognitive bias. In my reports I identify points at which I have made subjective decisions and inform the court that if they do not believe I have made an appropriate decision then the remainder of my analysis may be invalid.

9.7 In 1993, the US Supreme Court ruled in Daubert v Merrell Dow Pharmaceuticals [1993, 509 US 579] that when considering the admissibility of scientific evidence, the judge must consider the methodology's scientific validity and evidentiary reliability, including whether it has been empirically tested and found to have an acceptable error rate. In 2003 in United States v Robert N Angleton [2003, 269 F Supp 2nd 892 S D TX] the court conducted a relatively thorough review of the admissibility of the aural-spectrographic approach under US Federal Rule of Evidence 702 and case law following Daubert (the current version of Rule 702 includes requirements that "the testimony is based on sufficient facts or data" and "the testimony is the product of reliable principles and methods"). The court in Angleton ruled that the aural-spectrographic approach was not admissible. To my knowledge, the aural-spectrographic approach has not been accepted in US Federal Court since then (see summaries of rulings in Solan and Tiersma, 2003 pp 412–426; Faigman et al., 2008 §37.3, and Morrison, 2014 p 249).

9.8 In theory the validity and reliability of any system for the evaluation of forensic evidence could be empirically tested under conditions reflecting those of the case under investigation. This would apply equally to a system based on relevant data, quantitative measurements, and statistical models as to a system which is directly a human's subjective judgement based on their auditory and visual perception. In order

to test the validity and reliability of the system, a tester would present the system with a large number of pairs of samples, some of which the tester knows to be same-origin samples (two voice recordings from the same speaker) and some of which the tester knows to be different-origin samples (one voice recording from one speaker and one voice recording from another speaker). It is essential that the system being tested does not know the truth as to whether a test sample is a same-origin or a different-origin sample. In order for the results of the test to be relevant to the case under investigation, the recording conditions of the test samples must reflect the conditions of the audio recordings in the case and must be drawn from the relevant population. For each test pair, the system must output a quantification of the strength of evidence. The tester then compares the strength of evidence output by the system for each test pair with the truth as to whether that test pair was a same-origin pair or a different-origin pair. The tester then averages over the performance of the system on all the test pairs to derive a quantification of the validity of the system. The process for empirically evaluating the validity of a likelihood ratio system is described in detail in Morrison (2011), which also describes how to empirically evaluate the reliability of such a system.

9.9 Mr Primeau's report does not present any results of testing of the validity and reliability of his performance in applying the aural-spectrographic approach under conditions reflecting those of the case under investigation (or under any conditions). Without such testing, we do not know how good Mr Primeau is at determining the strength of the evidence under the conditions of this case (or under any conditions).

9.10 Mr Primeau states that he used "critical listening skills that [he] developed over the last 30 years". Experience may assist someone to get better at a task, but someone could also perform poorly at a task and not get any better irrespective of the amount of experience gained. In another matter, Stephanie Muse v Supervalue Inc [2011, United States District Court for the District of Maryland, 8:10-cv-01105-WGC], the court did not find Mr Primeau's then 25 years experience sufficient justification to admit his evidence: "Mr. Primeau's opinions are not reliable because they are not supported by adequate validation to render them trustworthy. Moreover, under Federal Rule of

Evidence 702, Mr. Primeau's proffered expert opinion testimony (1) is not based on sufficient facts or data and (2) is not the product of reliable principles and methods." (document 40, page 18).

9.11 Mr Primeau made the following observations:

> "• In all speech samples the male speaking has a pacing that is very significant and identifiable. It is quick and deliberate.
>
> • The voice tone in all three samples is identical.
>
> • When looking at the samples back to back in the comparison file, the wave formation and the sound spectrum are very similar, indicating a high percentage of certainty that these voices are from the same person."

9.12 Phonetics is the field of study concerned with speech sounds. "Pacing" is not to my knowledge a recognised term in phonetics. Mr Primeau does not define the term but from its everyday meaning I would surmise that it relates to speaking rate, i.e., how fast a person is speaking. There are quantitative metrics of speaking rate, but Mr Primeau does not provide any quantitative measurements of speaking rate related to the speech on the audio recording in this case. He qualitatively describes the speaking rate as "quick and deliberate". It is not clear whether he means this to refer to the portions of the recordings which are undisputedly of Mr Reynolds, the portions of the recordings which are undisputedly of Mr McTavish, or the portion of the recording in which the identity of the speaker is disputed. If he means this to apply to all three, then the probability of this evidence would be equally likely whether the disputed speaker was Mr Reynolds or whether it was Mr McTavish. If he means this to apply only to the portions of the recordings which are undisputedly of Mr Reynolds and the portion of the recording in which the identity of the speaker is disputed, then (in the terminology used above §8.3) he has only considered the similarity of the offender recording with respect to the suspect, he has not considered the typicality of the offender recording with respect to the relevant population. As described above (§8.3), similarity alone is not adequate for quantifying the strength of evidence.

9.13 "Tone" is a recognised term in phonetics, its quantitative acoustic correlate is fundamental frequency. Mr Primeau does not provide any quantitative measurements to support his claim that the fundamental frequency is identical. It is again not clear which portions of the recording Mr Primeau is referring to (one of the three samples could be the YouTube video he referred to, §9.1), so even taking his subjective opinion at face value the strength of the evidence is again either equal given the two hypotheses or does not take account of typicality. The term "identical" is also problematic. A person's fundamental frequency is highly variable and one would be highly unlikely to obtain exactly the same measured values for the fundamental frequency in each recording (say, mean fundamental frequency over the duration of each recording) even if both recordings were of the same speaker.

9.14 On the basis of visual inspection of graphical representations of the audio samples, Mr Primeau is of the opinion that they are very similar. Even if we accept Mr Primeau's subjective judgement in this matter, the fact that two things are similar does not, contrary to Mr Primeau's claim, imply that there is a high probability that they come from the same source. Even if the acoustic properties of two voice recordings are very similar, this alone does not imply that it is probable that they were produced by the same speaker, in the same way that if the offender has blond hair and the suspect has blond hair this alone does not imply that there is a high probability that the suspect and offender are the same person – typicality with respect to the relevant population must also be considered.

9.15 Mr Primeau concludes that he is "100% sure that the male at the beginning of the digital audio recording, and the male speaking at the one minute and fifty eight second mark, is the same person."

9.16 One could simply interpret Mr Primeau's conclusion as his opinion and disagree with his opinion without invoking any logical problems. There are, however, logical problems if one takes Mr Primeau's conclusion at face value and interprets it as a statement of the strength of evidence.

9.17 First, Mr Primeau's conclusion is an expression of the probability of the prosecution hypothesis given the evidence, not an expression of the probability of the evidence given the prosecution versus the defence hypotheses (the evidence being the acoustic properties of the voice on the second part of the recording, the prosecution hypothesis being that Dr Raynolds was the speaker on that part of the recording, and the defence hypothesis being that Mr McTavish was the speaker on that part of the recording). In order to draw a conclusion about the probability of the prosecution hypothesis given the evidence, one must consider additional information not related to the specific audio recordings which it is a forensic scientist's task to analyse. In order to draw a conclusion about the probability of the prosecution hypothesis versus the probability of the defence hypothesis given the evidence (posterior odds), one must consider not only the probability of the evidence given the prosecution hypothesis versus given the defence hypothesis (likelihood ratio), one must also consider ones belief as to the probability of the prosecution versus the defence hypothesis before the evidence was presented (prior odds). Formally, in order to calculate the posterior odds one must multiply the prior odds by the likelihood ratio. The prior odds reflect the trier of fact's belief about the strength of the competing hypotheses prior to the forensic scientist presenting their evidence (the trier of fact is the person or group of persons, a judge or a jury depending on the legal system, whose task it is to weigh all the evidence presented during the trial and decide on a verdict). The likelihood ratio presented by the forensic scientist as their estimate of the strength of the particular piece of evidence they have been asked to evaluate tells the trier of fact by how much the trier of fact should change their degree of belief as to the relative probability that the prosecution hypothesis is true versus that the defence hypothesis is true. It tells them the relative amount that their belief should change from before they hear the evidence to after they heard the evidence. The trier of fact's belief as to the relative probabilities of the two competing hypotheses before the evidence is presented may be influenced by other evidence which has already been presented in the case. Logically, if the forensic scientist presents posterior odds, then they must have substituted their own prior odds for those of the trier of fact and in doing so they have exceeded their rôle and intruded

into the rôle of the trier of fact.

9.18 The forensic scientist should not be informed as to the other evidence in the case since this could potentially have a biassing effect on their estimate of the strength of the particular piece of evidence that it is their task to evaluate. For example, if prior to evaluating fingerprint evidence the fingerprint examiner is told that the likelihood ratio from DNA evidence is in the billions and all that they have to do is corroborate the DNA evidence, then this will potentially bias the fingerprint examiner's estimate of the strength of the fingerprint evidence. Even before starting their analysis the fingerprint examiner will have a high belief that the prosecution hypothesis is true. High prior odds in favour of the prosecution hypothesis will bias the fingerprint examiner's posterior odds in favour of the prosecution hypothesis. In other scenarios, bias in favour of the defence hypothesis could occur. The forensic scientist should not consider other evidence in the case, they should only evaluate the strength of the particular evidence they have been asked to evaluate. Logically, the forensic scientist should produce a likelihood ratio, not a posterior probability. Even if a forensic scientist produces a likelihood ratio rather than posterior odds, if their decision as to the value of the likelihood ratio is based directly on their subjective judgement then they would be potentially highly susceptible to biassing information. As much as possible they should be blinded to potentially biassing information. Note that Mr Primeau's conclusion as to the strength of evidence was both an expression of posterior probability and based directly on his subjective judgement.

9.19 Second, Mr Primeau concludes that the posterior probability is 100% that the person speaking on the second part of the recording is the same speaker as in the first part of the recording (i.e., Dr Raynolds). Logically, a posterior probability of 100% can only be derived in two ways: either the prior probability is 100%, or the likelihood ratio is infinite (or both of these are true). If the prior probability is 100% then the strength of the evidence is irrelevant, the posterior probability will be 100% whatever the value of the likelihood ratio. If the likelihood ratio is infinite, then it does not matter what the prior odds were and it does not matter what the strength of any other evidence

relative to the hypotheses is, the posterior probability will be 100%. A likelihood ratio of infinity implies that no matter what other evidence is presented the prosecution hypothesis is true, no amount of alibi or eye witness evidence and no amount of other forensic evidence can outweigh this evidence. In forensic DNA analysis likelihood ratios can be extremely large, in the tens of billions or larger, but they are never infinite, they can always potentially be outweighed by other evidence. If the likelihood ratio from DNA evidence is very large, but the suspect has a cast iron alibi, then the trier of fact can decide that the strength of the alibi evidence outweighs that of the DNA evidence. If the likelihood ratio is infinite, and the trier of fact accepts this value, then the trier of fact cannot logically conclude that any other evidence could outweigh this evidence, and the forensic scientist has usurped the trier of fact's decision making rôle.

9.20 The structure of DNA data allows for potentially very large likelihood ratios to be calculated. This is not the case for speech data. In contrast to DNA profiles, the acoustic properties of speakers' voices are intrinsically highly variable from moment to moment, day to day, and year to year and are also highly susceptible to alteration due to different recording and transmission systems. In general, this results in a lot of overlap in the acoustic properties of recordings of different speakers. On average, this results in the potential size of likelihood ratios from voice evidence being much much smaller than the potential size of likelihood ratios from DNA evidence. Under typical casework conditions for forensic voice comparison I would not expect to obtain likelihood ratio values greater than in the thousands, and depending on the circumstances of the case will think we have done well to obtain values of around 10. In my opinion, it is entirely unjustified to make a claim of having obtained an infinite likelihood ratio from a forensic voice comparison.

10 Conclusion

10.1 Mr Primeau conducted a forensic voice comparison by listening to audio recordings

and looking at graphical representations of the acoustic properties of the audio recordings. This is an approach known as the aural-spectrographic approach. The practitioner uses aural and visual perception, and their conclusion as to the strength of the evidence is based directly on their subjective judgement. This approach has long been criticised for being overly subjective and susceptible to bias, and for its degree of validity and reliability not having been empirically demonstrated under conditions reflecting casework conditions. It has been ruled inadmissible in US Federal Court.

10.2 Mr Primeau did not provide any results of testing which would demonstrate the degree of his validity and reliability in performing forensic voice comparison under the conditions of the case under investigation (or under any conditions). Without such testing we do not know the level of his competence on this task.

10.3 Mr Primeau's conclusion of 100% certainty is logically flawed. Mr Primeau only took account of similarity, he did not take account of typicality. His conclusion is akin to concluding that if both the offender and the suspect have blond hair then they must be the same person. It does not take account of the typicality of blond hair in the population, i.e., the proportion of other people in the population who also have blond hair.

10.4 DNA evidence does not produce 100% certainty, and given the differences in data structure between DNA and voice data the strength of evidence which can be potentially extracted from voice evidence will be much much smaller than can potentially be extracted from DNA evidence.

10.5 The correct way to evaluate the strength of forensic evidence in a forensic voice comparison is via the use of relevant data, quantitative measurements of the acoustic properties of the recordings, and statistical models to calculate a likelihood ratio. The likelihood ratio takes into consideration both the similarity of the offender recording with respect to the suspect and the typicality of the offender recording with respect to the relevant population. The performance of the system, its validity and reliability, must be empirically tested under conditions reflecting those of the case under

investigation.

11 Legal References

Daubert v Merrell Dow Pharmaceuticals [1993, 509 US 579]

Federal Court of Canada Code of Conduct for Expert Witnesses [Schedule (Rule 52.2) of Federal Court Rules SOR/98-106, current to 2014-11-11 and last amended 2013-08-08]

Federal Rules of Evidence [2013] (United States)

Stephanie Muse v Supervalue Inc [2011, United States District Court for the District of Maryland, 8:10-cv-01105-WGC]

United States v Robert N Angleton [2003, 269 F Supp 2nd 892 S D TX]

12 Other References

Aitken CGG, Roberts P, Jackson G (2010) *Fundamentals of probability and statistical evidence in criminal proceedings: Guidance for judges, lawyers, forensic scientists and expert witnesses*. London: Royals Statistical Society.

Aitken CGG, Taroni F (2004) *Statistics and the evaluation of forensic evidence for forensic scientist*. 2nd ed. Chichester, UK: Wiley.

Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, 49, 161–164. doi:10.1016/j.scijus.2009.07.004

Balding DJ (2005) *Weight-of-evidence for forensic DNA profiles*. Chichester, UK: Wiley. doi:10.1002/9780470867693

Berger CEH, Buckleton J, Champod C, Evett IW, Jackson G (2011a) Evidence evaluation: A response to the Court of Appeal judgment in R v T. *Science & Justice*, 51, 43–49. doi:10.1016/j.scijus.2011.03.005

Bryden J (2014 November 27) Conservatives double-down on hidden-recording caper on Liberal, hire expert audio analyst. *The National Post*. http://news.nationalpost.com/2014/11/27/conservatives-double-down-on-hidden-recording-caper-on-liberal-hire-expert-audio-analyst/

Dror IE, Stoel RD (2014) Cognitive forensics: Human cognition, contextual information and bias. In Bruinsma G, Weisburd D, (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp 253–356). New York: Springer.

Enzinger E, Morrison GS (2014) Mismatched distances from speakers to telephone in a forensic-voice-comparison case. Submitted for publication.

Evett IW, Aitken CGG, Berger CEH, Buckleton JS, Champod C, Curran JM, Dawid AP, Gill P, González-Rodríguez J, Jackson G, Kloosterman A, Lovelock T, Lucy D, Margot P, McKenna L, Meuwly D, Neumann C, Nic Daeid N, Nordgaard A, Puch-Solis R, Rasmusson B, Radmayne M, Roberts P, Robertson B, Roux C, Sjerps MJ, Taroni F, Tjin-A-Tsoi T, Vignaux GA, Willis SM, Zadora G (2011) Expressing evaluative opinions: A position statement. *Science & Justice*, 51, 1–2. doi:10.1016/j.scijus.2011.01.002 doi:10.1016/S1355-0306(98)72117-3

Faigman DL, Saks MJ, Sanders J, Cheng EK, Eds (2008) *Modern Scientific Evidence: The Law and Science of Expert Testimony*. Eagan, MN: Thomson Reuters/West.

Foreman LA, Champod C, Evett IW, Lambert JA, Pope S (2003) Interpreting DNA Evidence: A Review. *International Statistics Journal*, 71, 473–495. doi:10.1111/j.1751-5823.2003.tb00207.x

Grigoras C, Smith JM, Morrison GS, Enzinger E (2013) Forensic audio analysis – Review: 2010–2013. In NicDaéid N. (Ed.), *Proceedings of the 17th International Forensic Science Mangers' Symposium, Lyon* (pp. 612–637). Lyon, France: Interpol.

Gruber JS, Poza F (1995) Voicegram identification evidence. *American Jurisprudence Trials*, 54(1), §1–§133.

Morrison GS (2009) Forensic voice comparison and the paradigm shift. *Science & Justice*, 49, 298–308. doi:10.1016/j.scijus.2009.09.002

Morrison GS (2010) Forensic voice comparison. In Freckelton I, Selby H, Eds., *Expert Evidence* (Ch. 99). Sydney, Australia: Thomson Reuters. http://expert-evidence.forensic-voice-comparison.net/

Morrison GS (2011) Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, 51, 91–98. doi:10.1016/j.scijus.2011.03.002

Morrison GS (2012) The likelihood-ratio framework and forensic evidence in court: A response to R v T. *International Journal of Evidence and Proof*, 16, 1–29. doi:10.1350/ijep.2012.16.1.390

Morrison GS (2014) Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison. *Science & Justice*, 54, 245–256. doi:10.1016/j.scijus.2013.07.004

Morrison GS, Evett IW, Willis SM, Champod C, Grigoras C, Lindh J, Fenton N, Hepler A, Berger CEH, Buckleton JS, Thompson WC, González-Rodríguez J, Neumann C, Curran JM, Zhang C, Aitken CGG, Ramos D, Lucena-Molina JJ, Jackson G, Meuwly D, Robertson B, Vignaux GA (2012) *Response to Draft Australian Standard: DR AS 5388.3 Forensic analysis - Part 3 - Interpretation*. http://forensic-evaluation.net/australian-standards/#Morrison_et_al_2012

Morrison GS, Stoel RD (2014) Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models – a response to Lennard (2013) Fingerprint identification: How far have we come? *Australian Journal of Forensic Sciences*, 46, 282–292. doi:10.1080/00450618.2013.833648

National Research Council (1979) *On the theory and practice of voice identification*. Washington, DC: National Academies.

National Research Council (2009) *Strengthening forensic science in the United States: A path forwards*. Washington, DC: National Academies.

Redmayne M, Roberts P, Aitken CGG, Jackson G (2011) Forensic science evidence in question. *Criminal Law Review*, 5, 347–356.

Robertson B, Vignaux GA, Berger CEH (2011) Extending the confusion about Bayes. Modern Law Review, 74, 444–455. doi:10.1111/j.1468-2230.2011.00857.x

Robertson B, Vignaux GA (1995) *Interpreting evidence: Evaluating forensic science in the courtroom*. Chichester, UK: Wiley.

Rose P (2002) *Forensic speaker identification*. London: Taylor & Francis.

Solan LM, Tiersma PM (2003) Hearing voices: speaker identification in court. *Hastings Law Journal*, 54, 373–435.

Tosi O (1979) *Voice identification: Theory and legal applications*. Baltimore, MD: University Park.

APPENDIX A

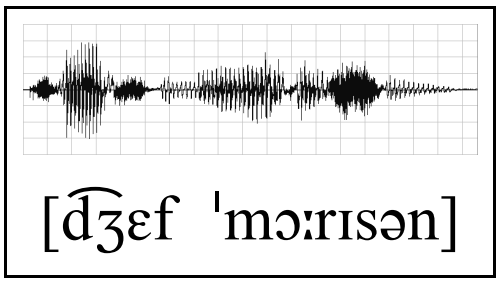
Curriculum Vitae

Geoffrey Stewart Morrison

Total 7 pages excluding this title page.

Geoffrey Stewart Morrison

BSc MTS MA PhD



Abridged Curriculum Vitae updated 30 November 2014

Contact Information

e-mail: [redacted]

websites: http://geoff-morrison.net/
http://forensic-evaluation.net/
http://forensic-voice-comparison.net/

telephone:
- Canada: [redacted]
- Australia: [redacted]

mailing address:
[redacted]
Vancouver BC
[redacted]
CANADA

Highlights

- **Adjunct Associate Professor**, Department of Linguistics, University of Alberta. 2010–present
- **Subject Editor**, *Speech Communication*. 2012–2014
- **Director**, Forensic Voice Comparison Laboratory, School of Electrical Engineering and Telecommunications, University of New South Wales. 2010–2013
- **Chair**, Forensic Acoustics Subcommittee, Acoustical Society of America. 2010–2013
- Total research funding brought in, 2010–2013: over $900k
- Total number of refereed and invited publications: 48
- Total number of forensic cases worked on: 17

Education

• PhD	Department of Linguistics, University of Alberta	2003 – 2006
• MA	Department of Linguistics, Simon Fraser University	2000 – 2002
• MTS	Vancouver School of Theology	1992 – 1995
• BSc	Faculty of Science & Engineering, University of Dundee	1998 – 1990

Research & Teaching Appointments

- Adjunct Associate Professor — 2010 – present
  Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada
- Visiting Fellow — 2013 – present
  School of Electrical Engineering and Telecommunications, University of New South Wales
- Senior Research Fellow / Director of Forensic Voice Comparison Laboratory — 2010 – 2013
  School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, New South Wales, Australia
- Visiting Fellow — 2009 – 2010
  School of Electrical Engineering and Telecommunications, University of New South Wales

• Invited Lecturer 2010 – 2013
Consejo Superior de Investigaciones Científicas (CSIC) [Spanish National Research Council],
/ Universidad Internacional Menéndez Pelayo (UIMP)
Judicial Phonetics Specialisation, Masters of Phonetics and Phonology Programme, Madrid, Spain

• Research Associate 2007 – 2010
School of Language Studies, Australian National University, Canberra, ACT, Australia

• Research Fellow 2006 – 2007
Speech Lab, Department of Cognitive & Neural Systems, Boston University, Boston, Massachusetts, USA

Competitive Research Funding

Grants

• Australian Research Council Linkage Project AU$ 544 013 2010–2013
Making demonstrably valid and reliable forensic voice comparison a practical everyday reality in Australia
Lead Investigator
with Julien Epps, Eliathamby Ambikairajah, Gary Edmond, Joaquín González-Rodríguez, Daniel Ramos, Cuiling Zhang
Partner Organisations: Australian Federal Police, National Institute of Forensic Science, Australasian Speech Science and Technology Association, New South Wales Police, Queensland Police Service, Guardia Civil

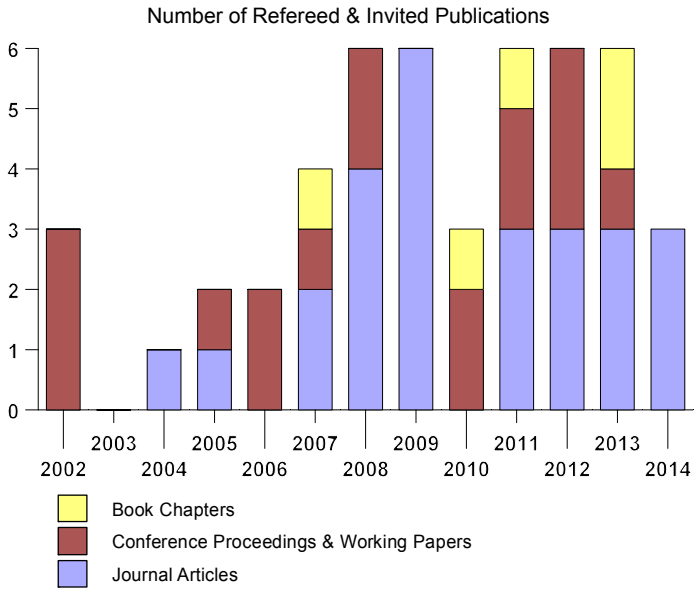
• United States Government, Office of the Director of National Intelligence (ONDI), US$confidential 2010–2011
Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL),
Subcontract through IBM Thomas J. Watson Research Center
Incorporation of forensic analysis techniques as part of an automatic speaker recognition system
Principal Investigator at IBM: Jason W. Pelecanos
Other IBM researches: Mohamed Omar, Weizhong Zhu, Sibel Yaman, Kyu Han
Principal Investigator at UNSW: Geoffrey Stewart Morrison
Other UNSW researchers:
Cuiling Zhang, Felipe Ochoa, Ewald Enzinger, Tharmarajah Thiruvaran, William Steed, Eugenia San Segundo Fernández

• Spanish Ministry of Science and Innovation, € 66 000 2010–2013
National Programme for Fundamental Research Projects
Individual voice quality in the identification of speakers
with Juana Gil Fernández (Lead Investigator), Secundino Fernández González,
María José Albalá Hernández, José Antonio Hierro, Jorge Rico Ródenas, Eugenia San Segundo

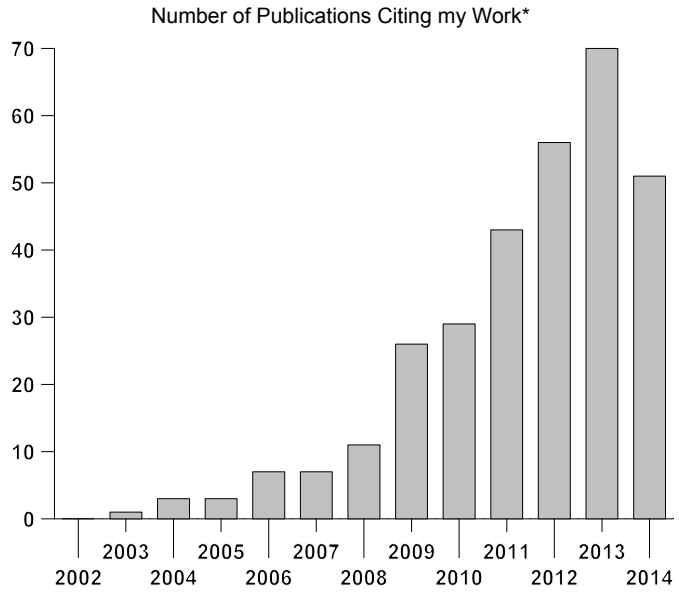
Awards

• Australian Research Council Postdoctoral Fellowship (Industry) AU$ 245 538 awarded 2010
• Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship CAN$ 81 000 awarded 2006
• Social Sciences and Humanities Research Council of Canada Doctoral Fellowship CAN$ 59 000 awarded 2003
• University of Alberta Honorary PhD Scholarship and Walter H Johns Graduate Fellowship CAN$ 14 566 awarded 2003

Publication Statistics



2014 counts may include in-press publications



*Excludes publications in which I cite my own work. Publications which cite two or more of my publications are counted only once. There may be additional publications citing my work but not included in these counts. Count for 2014 not complete.

Selected Publications

Morrison, G. S., & Stoel, R. D. (2014). Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models – a response to Lennard (2013) Fingerprint identification: How far have we come? *Australian Journal of Forensic Sciences, 46*, 282–292. doi:10.1080/00450618.2013.833648

Morrison, G. S. (2014). Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison. *Science & Justice*, 54, 245–256. doi:10.1016/j.scijus.2013.07.004

Morrison, G. S., Lindh, J., & Curran, J. M. (2014). Likelihood ratio calculation for a disputed-utterance analysis with limited available data. *Speech Communication*, 58, 81–90. doi:10.1016/j.specom.2013.11.004

Grigoras, C., Smith, J. M., Morrison, G. S., & Enzinger, E. (2013). Forensic audio analysis – Review: 2010–2013. In: NicDaéid, N. (Ed.), *Proceedings of the 17th International Forensic Science Mangers' Symposium, Lyon* (pp. 612–637). Lyon, France: Interpol.

Zhang, C., Morrison, G. S., Enzinger, E., & Ochoa, F. (2013). Effects of telephone transmission on the performance of formant-trajectory-based forensic voice comparison – female voices. *Speech Communication*, 55, 796–813. doi:10.1016/j.specom.2013.01.011

Zhang, C., Morrison, G. S., Ochoa, F., & Enzinger, E. (2013). Reliability of human-supervised formant-trajectory measurement for forensic voice comparison. *Journal of the Acoustical Society of America. 133*, EL54–EL60. doi:10.1121/1.4773223

Morrison, G. S. (2013). Tutorial on logistic-regression calibration and fusion: Converting a score to a likelihood ratio. *Australian Journal of Forensic Sciences*, 45, 173–197. doi:10.1080/00450618.2012.733025

Morrison, G. S. (2013). Vowel inherent spectral change in forensic voice comparison. In G. S. Morrison & P. F. Assmann (Eds.) *Vowel inherent spectral change* (ch. 11 / pp. 263–282). Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-14209-3_11

Enzinger, E. & Morrison, G. S. (2012). The importance of using between-session test data in evaluating the performance of forensic-voice-comparison systems. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney* (pp. 137–140). Australasian Speech Science and Technology Association.

Morrison, G. S., Ochoa, F., & Thiruvaran, T. (2012). Database selection for forensic voice comparison. In *Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore* (pp. 62–77). International Speech Communication Association.

Enzinger, E., Zhang, C., & Morrison, G. S. (2012). Voice source features for forensic voice comparison – an evaluation of the GLOTTEX® software package. In *Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore* (78–85). International Speech Communication Association. [errata and addenda available on my website]

Morrison, G. S. (2012). The likelihood-ratio framework and forensic evidence in court: A response to R v T. *International Journal of Evidence and Proof*, 16, 1–29. doi:10.1350/ijep.2012.16.1.390

Morrison, G. S., Rose, P., & Zhang, C. (2012). Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. *Australian Journal of Forensic Sciences*, 44, 155–167. doi:10.1080/00450618.2011.630412

Lindh, J., & Morrison, G. S. (2011). Forensic voice comparison by humans and machine: Forensic voice comparison on a small database of Swedish voice recordings. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, China* (pp. 1254–1257). Hong Kong: Organizers of ICPhS XVII at the Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Zhang, C., Morrison, G. S., & Thiruvaran, T. (2011). Forensic voice comparison using Chinese /iau/. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, China* (pp. 2280–2283). Hong Kong: Organizers of ICPhS XVII at the Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Morrison, G. S. (2011). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, 51, 91–98. doi:10.1016/j.scijus.2011.03.002

Morrison, G. S. (2011). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multvariate kernel density (MVKD) versus Gaussian mixture model – universal background model (GMM-UBM). *Speech Communication*, 53, 242–256.

Morrison, G. S., Zhang, C., Rose, P. (2011). An empirical estimate of the precision of likelihood ratios from a forensic-voice-comparison system. *Forensic Science International*, 208, 59–65. doi:10.1016/j.forsciint.2010.11.001

Morrison, G. S. (2010). Forensic voice comparison. In I. Freckelton, & H. Selby (Eds.), *Expert Evidence* (Ch. 99). Sydney, Australia: Thomson Reuters.

Morrison, G. S., Thiruvaran, T., Epps, J. (2010). Estimating the precision of the likelihood-ratio output of a forensic-voice-comparison system. *Proceedings of Odyssey 2010: The Language and Speaker Recognition Workshop*, Brno, Czech Republic (pp. 63–70).

Morrison, G. S. (2009). Forensic voice comparison and the paradigm shift. *Science & Justice, 49*, 298–308.

Morrison, G. S. (2009). Comments on Coulthard & Johnson's (2007) portrayal of the likelihood-ratio framework. *Australian Journal of Forensic Sciences, 41*, 155–161.

Rose, P., & Morrison, G. S. (2009). A response to the UK position statement on forensic speaker comparison. *International Journal of Speech, Language and the Law, 16*, 139–163.

Morrison, G. S. (2009). Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *Journal of the Acoustical Society of America, 125*, 2387– 2397.

Morrison, G. S., & Kinoshita, Y. (2008). Automatic-type calibration of traditionally derived likelihood ratios: Forensic analysis of Australian English /o/ formant trajectories. *Proceedings of Interspeech 2008 Incorporating SST 2008* (pp. 1501–1504). International Speech Communication Association.

Zhang, C., Morrison, G. S., & Rose, P. (2008). Forensic speaker recognition in Chinese: A multivariate likelihood ratio discrimination on /i/ and /y/. *Proceedings of Interspeech 2008 Incorporating SST 2008* (pp. 1937–1940). International Speech Communication Association.

Morrison, G. S. (2008). Forensic voice comparison using likelihood ratios based on polynomial curves fitted to the formant trajectories of Australian English /aɪ/. *International Journal of Speech, Language and the Law, 15*, 249–266.

Selected Workshops, Tutorials, and Presentations

Workshops and Talks at Forensic Laboratories

Morrison, G. S. (2013, May). *Workshop on forensic speech science*. Workshop given at Institut de Recherche Criminelle de la Gendarmerie Nationale, Paris, France.

Morrison, G. S. (2013, April). *Testing the validity and reliability of likelihood-ratio systems & Calibrating likelihood-ratio systems*. Talk given at the Netherlands Forensic Institute, The Hague, The Netherlands.

Morrison, G. S. (2013, April). *Introduction to the likelihood-ratio framework for the evaluation of forensic evidence*. Workshop given at Victoria Police Forensic Services Centre, Melbourne, Victoria, Australia.

Morrison, G. S. (2013, February). *Introduction to the likelihood-ratio framework for the evaluation of forensic evidence*. Workshop given at Queensland Police Forensic Services Branch, Brisbane, Queensland, Australia.

Morrison, G. S. (2010, April). *Forensic voice comparison and the paradigm shift in forensic science*. Talk given at the Netherlands Forensic Institute, The Hague, The Netherlands.

Morrison, G. S. (2009, July). *The place of forensic voice comparison in the ongoing paradigm shift*. Talk given at the Forensic and Data Centres, Australian Federal Police, Canberra, ACT, Australia.

Morrison, G. S. (2008, July). *Combining acoustic-phonetic and automatic approaches to forensic speaker comparison: Calibration and fusion of likelihood-ratios extracted from the formant trajectories of diphthongs*. Talk given at Sprecher-Erkennung und Tonträgeranalyse [Department of Speaker Identification and Audio Analysis], Bundeskriminalamt [German Federal Police], Wiesbaden, Germany.

Tutorials and Workshops at Major Conferences

Morrison, G. S. (2012, August). *An introduction to data-based calculation of likelihood ratios and assessment of validity and reliability*. Workshop presented at the European Academy of Forensic Science (EAFS) Conference, The Hague, The Netherlands.

Morrison, G. S. (2012, June). *Workshop on validity and reliability in forensic voice comparison*. Invited workshop presented at National Center for Media Forensics, University of Colorado Denver immediately prior to the 46th Audio Engineering Society (AES) Conference on Audio Forensics: Recording, Recovery, Analysis, and Interpretation, Denver, Colorado, USA.

Morrison, G. S., & Ramos, D. (2010, November). *Forensic voice comparison.* Tutorial at the 2nd Pan-American/Iberian Meeting on Acoustics / 160th Meeting of the Acoustical Society of America, Cancún, Quintana Roo, Mexico.

Morrison, G. S. (2010, June). *Tutorial on forensic comparison of audio recordings in the same framework as is standard for forensic comparison of DNA profiles*. Tutorial at the 39th International Audio Engineering Society (AES) Conference – Audio Forensics: Practices and Challenges, Hillerød, Denmark.

Kinoshita, Y., Morrison, G. S., & Ramos, D. (2008, September). *Forensic speaker comparison - Likelihood ratios - As not seen on TV*. Tutorial at the Interspeech 2008 Conference, Brisbane, Queensland, Australia.

Key Conference Presentations

Morrison, G. S., Enzinger, E. (2014, August). *Forensic likelihood ratios should not be based on similarity scores or difference scores*. Paper presented at the 9th International Conference on Forensic Inference and Statistics (ICFIS), Leiden, The Netherlands.

Morrison, G. S. (2013, September). *An introduction to the likelihood-ratio framework for the evaluation of forensic evidence*. Paper presented at the 36th Canadian Identification Society Annual Education Conference, Vancouver, British Columbia, Canada.

Morrison, G. S. (2013, June). *Distinguishing between science and pseudoscience in forensic acoustics*. Paper presented at the 21st International Congress on Acoustics, Montréal, Québec, Canada.

Morrison, G. S., (2012, June). *How does forensic voice comparison differ from automatic speaker recognition? With a particular focus on database selection*. Chinese and Oriental Languages Information Processing Society (COLIPS) Distinguished Lecturer at Nanyang Technological University (NTU) in conjunction with Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore.

Morrison, G. S., & Hoy, M. (2012, June). *What did Bain really say? A preliminary forensic analysis of the disputed utterance based on data, acoustic analysis, statistical models, calculation of likelihood ratios, and testing of validity*. Invited paper presented at the 46th Audio Engineering Society (AES) Conference on Audio Forensics: Recording, Recovery, Analysis, and Interpretation, Denver, Colorado, USA.

Morrison, G. S. (2011, February). *The new paradigm in forensic science*. Invited presentation at the National Judicial College of Australia Expert Evidence Conference, Canberra, Australian Capital Territory, Australia.

Morrison, G. S., (2010, October). *Measuring validity and reliability in forensic science*. Keynote presentation at BIT's 1st Annual World Congress of Forensic Science, Dalian, Liaoning, China.

Morrison, G. S., (2009, July). *The place of forensic voice comparison in the ongoing paradigm shift*. Invited presentation at the 2nd International Conference on Evidence Law and Forensic Science, Beijing, China.

Forensic Casework

- worked on a total of 17 cases
- written reports submitted in relation to 14 cases (7 at the behest of the prosecution/plaintiff and 7 at the behest of the defence/respondent)
- contribution to 2 amicus briefs
- oral testimony in court in 4 cases (all at the behest of the defence/respondent)

- Supreme Court of the United States 2013
  Clacy Watson Herrera v United States, No. 12-1461
  - One of 25 scientists and scholars contributing to an Amicus Brief.

• Crime Investigation Unit, Victoria Police, Boroondara, VIC, Australia 2012–2013
- Preliminary report on forensic voice comparison.
- Forensic voice comparison - report on full analysis submitted - suspect changed plea

• Emery Partners Solicitors, Newcastle, NSW, Australia 2012–2013
- Forensic voice comparison - report on full analysis submitted

• Aquila Lawyers, Sydney, NSW, Australia 2012
R v Christina My Phung Ly
- Report on preliminary analysis of voice recording and critique of a forensic-voice-comparison report produced by another forensic scientist.
- Presentation of oral evidence in court (voir dire and before jury).

• Fisher Dore Lawyers, Brisbane, QLD, Australia 2012
Peter Foster ats Australian Competition and Consumer Commission
- Report on preliminary analysis of voice recording and critique of a forensic-voice-comparison report produced by another forensic scientist.
- Presentation of oral evidence in court.

• Herbert Geer Lawyers, Melbourne, VIC, Australia 2012
- Preliminary report on forensic voice comparison.

• South Australian Office of the Director of Public Prosecutions, Adelaide, SA, Australia, 2012
and Criminal Investigations Branch, South Australia Police, Holden Hill, SA, Australia
- Written critique of a forensic-voice-comparison report produced by another forensic scientist.

• Criminal Investigations Branch, South Australia Police, Port Augusta, SA, Australia 2012
- Preliminary report on forensic voice comparison.

• United States Court of Appeals for the Ninth Circuit 2012
Nelson Acosta-Roque v Eric Holder Jr, No. 11-70705
- One of 39 scientists and scholars contributing to an Amicus Brief.

• Henry Sklarz Lawyers, Perth, WA, Australia 2011
State of Western Australia v Thi Dieu Linh Lai [WA Dist Ct, No 654 of 2011]
- Presentation of oral evidence in court on non-technical speaker identification by lay persons.

• Garde-Wilson Lawyers, Melbourne, VIC, Australia 2009
- Preliminary report on forensic voice comparison.

• D G Price & Co, Barristers & Solicitors, Perth, WA, Australia 2009
State of Western Australia v Cameron James Mansell [WA Dist Ct, No 665 of 2008]
- Written report on non-technical speaker identification by lay persons submitted to court.
- Presentation of oral evidence in court.

• Purana Taskforce, Victoria Police, Melbourne, VIC, Australia 2009
- Preliminary report on forensic voice comparison

• South East Asian Crime Squad, New South Wales Police, Sydney, NSW, Australia 2009
- Preliminary report on forensic voice comparison

• Jim Young, Barrister-at-Law, Sydney, NSW, Australia 2009
- Written report on non-technical speaker identification by lay persons submitted to court.

• Ford Criminal Lawyers, Sydney, NSW, Australia 2008
- Preliminary report on forensic voice comparison

• Major Crash Investigation Unit, South Australia Police, Adelaide, SA, Australia 2008
- Preliminary report on forensic voice comparison

Other Activities

Journals

• Speech Communication
Subject Editor, including responsibility for papers on forensic speech science 2012 –2014

• Reviewer
Science & Justice, Forensic Science International, Australian Journal of Forensic Sciences, Sydney Law Review, Journal of the Acoustical Society of America, Journal of Phonetics

Conference Organising

• Organiser with Joseph Campbell of special session: *Forensic Voice Comparison and Forensic Acoustics – Distinguishing between science and pseudoscience in forensic acoustics* June 2013
21st International Congress on Acoustics (Montréal, Québec, Canada)

• Organiser of special session: *Forensic Voice Comparison and Forensic Acoustics – On the leading edge of the tidal wave of change about to hit forensic science in the US(?)* November 2011
162nd Meeting of the Acoustical Society of America (San Diego, California, USA)

• Organiser of special session: *Forensic Voice Comparison and Forensic Acoustics* November 2010
2nd Pan-American/Iberian Meeting on Acoustics (Cancún, Quintana Roo, Mexico)

• Organiser of special session: *Forensic Speaker Recognition Traditional and Automatic Approaches* September 2008
Interspeech 2008 (Brisbane, Queensland, Australia)

Associations

• Acoustical Society of America,
Chair of the Forensic Acoustics Subcommittee 2010 – 2013

• Australasian Speech Science & Technology Association
Member of Forensic Speech Science Committee 2010 – 2013

• International Association for Forensic Phonetics and Acoustics
Member of Research Committee 2010 – 2012

APPENDIX B

Federal Court of Canada Code of Conduct for Expert Witnesses

Total 2 pages excluding this title page.

SCHEDULE
(Rule 52.2)

CODE OF CONDUCT FOR EXPERT WITNESSES

General Duty to the Court

1. An expert witness named to provide a report for use as evidence, or to testify in a proceeding, has an overriding duty to assist the Court impartially on matters relevant to his or her area of expertise.

2. This duty overrides any duty to a party to the proceeding, including the person retaining the expert witness. An expert is to be independent and objective. An expert is not an advocate for a party.

Experts' Reports

3. An expert's report submitted as an affidavit or statement referred to in rule 52.2 of the *Federal Courts Rules* shall include

(*a*) a statement of the issues addressed in the report;

(*b*) a description of the qualifications of the expert on the issues addressed in the report;

(*c*) the expert's current *curriculum vitae* attached to the report as a schedule;

(*d*) the facts and assumptions on which the opinions in the report are based; in that regard, a letter of instructions, if any, may be attached to the report as a schedule;

(*e*) a summary of the opinions expressed;

(*f*) in the case of a report that is provided in response to another expert's report, an indication of the points of agreement and of disagreement with the other expert's opinions;

(*g*) the reasons for each opinion expressed;

(*h*) any literature or other materials specifically relied on in support of the opinions;

(*i*) a summary of the methodology used, including any examinations, tests or other investigations on which the expert has relied, including details of the qualifications of the person who carried them out, and whether a representative of any other party was present;

(*j*) any caveats or qualifications necessary to render the report complete and accurate, including those relating to any insufficiency of data or research and an indication of any matters that fall outside the expert's field of expertise; and

(*k*) particulars of any aspect of the expert's relationship with a party to the proceeding or the subject matter of his or her proposed evidence that might affect his or her duty to the Court.

4. An expert witness must report without delay to persons in receipt of the report any material changes affecting the expert's qualifications or the opinions expressed or the data contained in the report.

ANNEXE
(règle 52.2)

CODE DE DÉONTOLOGIE RÉGISSANT LES TÉMOINS EXPERTS

Devoir général envers la Cour

1. Le témoin expert désigné pour produire un rapport qui sera présenté en preuve ou pour témoigner dans une instance a l'obligation primordiale d'aider la Cour avec impartialité quant aux questions qui relèvent de son domaine de compétence.

2. Cette obligation l'emporte sur toute autre qu'il a envers une partie à l'instance notamment envers la personne qui retient ses services. Le témoin expert se doit d'être indépendant et objectif. Il ne doit pas plaider le point vue d'une partie.

Les rapports d'expert

3. Le rapport d'expert, déposé sous forme d'un affidavit ou d'une déclaration visé à la règle 52.2 des *Règles des Cours fédérales*, comprend :

a) un énoncé des questions traitées;

b) une description des compétences de l'expert quant aux questions traitées;

c) un *curriculum vitae* récent du témoin expert en annexe;

d) les faits et les hypothèses sur lesquels les opinions sont fondées, et à cet égard, une lettre d'instruction peut être annexée;

e) un résumé des opinions exprimées;

f) dans le cas du rapport qui est produit en réponse au rapport d'un autre expert, une mention des points sur lesquels les deux experts sont en accord et en désaccord;

g) les motifs de chacune des opinions exprimées;

h) les ouvrages ou les documents expressément invoqués à l'appui des opinions;

i) un résumé de la méthode utilisée, notamment des examens, des vérifications ou autres enquêtes sur lesquels l'expert se fonde, des détails sur les qualifications de la personne qui les a effectués et une mention quant à savoir si un représentant des autres parties était présent;

j) les mises en garde ou réserves nécessaires pour rendre le rapport complet et précis, notamment celles qui ont trait à une insuffisance de données ou de recherches et la mention des questions qui ne relèvent pas du domaine de compétence de l'expert;

k) tout élément portant sur la relation de l'expert avec les parties à l'instance ou le domaine de son expertise qui pourrait influencer sur son devoir envers la Cour.

4. Le témoin expert doit signaler immédiatement aux personnes qui ont reçu le rapport tout changement important ayant une incidence sur ses qualifications et les opinions exprimées ou sur les données figurant dans le rapport.

Expert Conferences

5. An expert witness who is ordered by the Court to confer with another expert witness

(*a*) must exercise independent, impartial and objective judgment on the issues addressed; and

(*b*) must endeavour to clarify with the other expert witness the points on which they agree and the points on which their views differ.

SOR/2010-176, s. 13.

Conférences d'expert

5. Le témoin expert à qui la Cour ordonne de s'entretenir avec un autre témoin expert doit, à la fois ;

a) faire preuve d'un jugement indépendant, impartial et objectif quant aux questions traitées;

b) s'efforcer de clarifier avec les autres témoins experts les points sur lesquels ils sont en accord et ceux sur lesquels ils ont une divergences d'opinions.

DORS/2010-176, art. 13.

APPENDIX C

International Association for Forensic Phonetics and Acoustics Code of Practice

Total 1 page excluding this title page.

International Association for Forensic Phonetics and Acoustics Code of Practice

http://www.iafpa.net/code.htm

This is the current IAFPA Code of Practice which was approved by the AGM in Helsinki, 2004.

1. Members should act in all circumstances with integrity, fairness and impartiality.

2. Recognising the varied array of casework subsumed under the interests of IAFPA (eg speaker identification/elimination, speaker profiling, voice line-ups, transcription, authentication, signal enhancement, sound propagation at crime scenes), Members should maintain awareness of the limits of their knowledge and competencies when agreeing to carry out work.

3. Members should not enter into any arrangements in which remuneration is dependent on the outcome of the case.

4. Members should make clear, both in their reports and in giving evidence in court, the limitations of forensic phonetic and acoustic analysis.

5. In reporting on cases where an opinion or conclusion is required, Members should make clear their level of certainty and give an indication of where their conclusion lies in relation to the range of judgements they are prepared to give.

6. (a) Members should exercise particular caution if carrying out forensic analysis of any kind on recordings containing speech in languages of which they are not native speakers.

   (b) In carrying out forensic speaker identification/elimination work, Members should exercise particular caution if the samples for comparison are in different languages.

   (c) Members should exercise particular caution if carrying out authenticity or integrity examinations of recordings that are not claimed to be original.

7. Members undertaking forensic phonetic and acoustic analyses or operations of all kinds should state in their reports the methods they have followed and provide details of the equipment and computer programs used.

8. Members, in making their analysis, should take due account of the methods available at the time and of their appropriateness to the samples under examination.

9. Members should not attempt to do psychological profiles or assessments of the sincerity of speakers.

10. Member's reports should not include or exclude any material which has been suggested by others (in particular by those instructing them) unless that Member has formed an independent view.

APPENDIX D

Primeau Report

Total 3 pages excluding this title page.

PRIMEAU FORENSICS

IDENTIFY | CLARIFY | TESTIFY

26 November 2014

Michael Oberman

Dear Mr. Oberman,

I am an audio and video forensic expert and have been practicing for over 30 years. I have testified in several courts throughout the United States and worked on various international cases. My forensic practices for audio investigation include digital and analogue audio authentication, restoration and voice identification. As a video forensic expert, my practices include video authentication, restoration and identification.

On November 25, 2014, you sent me an email which contained a digital audio recording file, titled 'Audio Recording.MP4.' The recording is of an interview between a male and a female. You asked that I determine if the male speaking in the beginning of the recording matched the voice of the male speaking after the one minute and fifty eight second mark. In that same email you included a link to a YouTube video purportedly featuring Marlo Raynolds. You also asked for me to determine if the male in the YouTube video is the same male that is speaking in both portions of the digital audio recording.

1878 Star Batt Drive, Bldg. 2E Rochester Hills, MI 48309 800-647-4281 Fax: 248-289-1869

The goal of a voice identification test is to compare the known and unknown voices using critical listening, electronic measurement and visual inspection of sound wave formation and color sonogram.

This report will include descriptions of the similarities observed during critical listening, electronic measurement and visual inspection testing.

I began my voice identification testing by loading the digital audio file in to my forensic computer using Adobe Audition. I removed the audio file from the YouTube video as well, also using Adobe Audition. Using critical listening skills that I developed over the last 30 years I listened to the known and unknown voices repeatedly until I was familiar with all voice recordings.

I then created a composition audio file placing speech samples back to back so that I could critically listen to the similarities as well as the differences.

I made the following observations:

- In all speech samples the male speaking has a pacing that is very significant and identifiable. It is quick and deliberate.
- The voice tone in all three samples is identical.
- When looking at the samples back to back in the comparison file, the wave formation and the sound spectrum are very similar, indicating a high percentage of certainty that these voices are from the same person.

Please note that the words that I compared to arrive at my conclusion were not exact. In other words, the voice identification test was conducted using different speech samples.

I am 100% sure that the male at the beginning of the digital audio recording, and the male speaking at the one minute and fifty eight second mark, is the same person. I am 75-80% sure that the male speaking in the digital audio recording is the same person that is speaking in the YouTube video.

In order to further pursue voice identification testing I would require an exemplar from the suspected individual saying the exact words that are spoken in the unknown portion of the digital audio recording.

This concludes this voice identification report. Please let me know if I can assist further with this investigation.

Respectfully submitted,

Edward J Primeau CCI, CFC

1878 Star Batt Drive, Bldg. 2E Rochester Hills, MI 48309 800-647-4281 Fax: 248-289-1869