

Quantitative Comparison Method of Forensic Voice Evidence

Huapeng W*

Department of Forensic Science and Technology, Criminal Investigation Police
University of China, Shenyang, 110854, China

Research Article

Volume 2 Issue 2

Received Date: September 22, 2017

Published Date: October 24, 2017

***Corresponding author:** Wang Huapeng, Key Laboratory of Evidence Science, Ministry of Education, Beijing, 10088, Num 83, Ta wan street of Huang gu district, Shenyang, 110854, China, Tel: 086-02486982220; Email: huapeng.wang@hotmail.com

Abstract

In this paper, the scientific demands of forensic evidence were introduced. The quantitative analysis methods and processes of forensic evidence were proposed, and the feasible quantitative examination methods of evidence that match the admission criterion of scientific evidence were summarized. In addition, taking voice evidence for example, the data-driven calculation methods of evidence strength based on statistical analysis were provided, which can also be applied to other forensic fields.

Keywords: Evidence quantification; Evidence strength; Scientific evidence

The Demands of Scientific Evidence

For forensic science, establishing an evidence evaluation framework that matches logic and follows scientific principle is essential. Forensic scientist can provide effective information to courts through this framework to avoid providing confusing and unclear conclusion. Although a systematic study on forensic recognition had been carried out in the 20th century, especially in the fields of gunshot, fingerprint, glass, tool marks and footprint [1-6], how to provide scientific evidence for courts is still a hot debate in many forum of law and forensic science. One reason is that the U.S. Supreme Court promulgate Daubert rulings [7] about the admissibility of evidence in 1993. In accordance with the rulings, the U.S. Supreme Court suggested that the forensic evidence identification technology must have a standard procedure, and can be demonstrated its test performance and accuracy, and accepted by the scientific community at the same time. If examination results are the unscientific opinions, such as the expert evidence that

lack of scientific basis, will not be adopted by the court. These guidelines consistent with the opinions of many forensic scientists in the world, which requires a more transparent principle and a logical theory to interpret forensic evidence. In Daubert rulings specified in the terms, is to ensure that the examination results are through scientific theories, reasoning or methodology that is reliable and valid. These conditions can be summarized as follows [8]:

- Whether the theory used can be tested and has been tested.
- Whether the technique used has been published or subjected to peer review.
- Whether the technique used has a known or potential rate of error in application.
- Whether the relevant standards exist and are maintained to ensure the operation of the technique.
- Whether the technique is generally accepted in the relevant scientific community.
- Whether the technique is based on the data of a type reasonably relied on by experts in the field or facts.

•Whether the technique has a quantitative probative value and not outweighed by the dangers of incorrect prejudice, confusion of issues or misleading the jury.

Daubert rulings require that examination methods of forensic evidence must have been tested, and error ratios must have been known and adopted, which are also the kernel requirement of scientificity and reliability in evidence evaluation. Most countries in the world have realized the significance of this issue. In February 2009, the United States national research council's report submitted to congress (NRC, 2009), and in April 2010, England and Wales Law Commission consultation paper (Law themselves of England and Wales, 2009) were both strongly calls attention to the reliability of forensic science. Scientific, accuracy and objectivity of forensic evidence technology were put forward the urgent request. Due to the stability and specificity of different kinds of evidence itself, examination technology, methods and standards also vary. This can make all kinds of evidence out of balance between evidence strength and reliability degree. DNA, for example, it has better stability and specificity than voice, handwriting and footprint, and therefore, evidence strength and reliability in general better than the latter, this is decided by different attributes and characteristics of evidence itself. Reading Daubert standards, its requirement is that regardless of any evidence, as evidence in court when using, should shows evidence strength and the accuracy and reliability of the method of examination technology, and has clear examination principle and quantifiable, repeatable and objective examination result.

Quantitative method of evidence characteristic

To realize the quantitative analysis of evidence, first of all, the characteristics of evidence must be expressed in numerical. Only after quantification of the characteristics, a quantitative identification results can be calculated. The characteristics and the method of digitalization indifferent domains of forensic science are very different. In this paper, taking the voice evidence for example, we describe the digitalization methods of characteristics.

In the evaluation of voice evidence, the traditional voiceprint identification method concludes the auditory examination and visual voiceprint examination, but both are based on the experience of the subjective judgment of expert, and unable to provide precise and repeatable quantitative conclusion. To achieve the quantitative conclusion of voice evidence, the first step is to realize quantitative representation of the phonetic

characteristics. Voice features are usually divided into acoustics-phonetics features and digital features. Commonly used acoustic- phonetics features include fundamental frequency, all levels of formant frequency and bandwidth, etc. Such features are generally extracted manually through the Praat or smart voiceprint identification workstation software. In the process of extraction, examiner's knowledge about formants and fundamental frequency will influence the results, and even the same examiner is difficult to get the same results in two operations (because this operation uses the visual inspection of mouse), so it does not meet the requirements of precision repeatability. Such features can also use computer to extract automatically, but in large or great change noise speech evidence, the accuracy of computer extraction is less than artificial extraction, therefore, the automatic method are often be abandoned. With the development of acoustic -phonetic feature extraction algorithm, the results of computer is becoming more and more accurate, especially in the vowel parts, due to the vibration energy is bigger, and the automatic method can provide relatively accurate extraction results. Meanwhile, vowel sections contain the most important distinguishable information in speaker identity. Therefore, if we use acoustic-phonetic features, we can retain vowel part data and remove the data of consonant part. Another kind of features are the digital features which computer automatically extracted, such as linear prediction coefficients, linear prediction cepstral coefficient, Mel - frequency cepstral coefficients, partial correlation coefficient, short-time zero crossing ratio, and short-term spectrum, etc. Since those feature are discrete values, therefore, meet the requirements of the quantitative analysis in evidence evaluation.

Data-Driven Calculation Methods of Evidence Strength Based on Statistical Analysis

After the quantification of features, data-driven calculation methods of evidence strength based on statistical analysis can be used to evaluate the strength of evidence. Complete and high quality evidence can provide stronger evidence strength, however, a small amount, incomplete or contaminated evidence can only provide less strong evidence strength. The above criteria are consistent with objective reality, and are a kind of evaluation method match logic reasoning. Internationally at present, the likelihood ratio (LR) evaluation system of evidence meets the requirements in Daubert rulings, and has been used in glass, paint, handwriting, DNA, voice, and other fields of evidence identification, more

important of all, the method has been accepted by the relevant scientific community [9-11]. According to different types of evidence, the calculation method is a little different, here are two methods used widely in evidence evaluation.

Idiot's Bayes Method

In 1977, Lindley proposed a calculation method of likelihood ratio [12], which is suitable in the case of single variable or computes multiple variables one by one. In Lindley's formula, the seven parameters are considered, see formula (1).

$$v \cong \frac{\tau}{a\sigma} \times \exp \left\{ -\frac{(\bar{x} - \bar{y})^2}{2a^2\sigma^2} \right\} \times \exp \left\{ -\frac{(w - \mu)^2}{2\tau^2} + \frac{(z - \mu)^2}{\tau^2} \right\}$$

In which, \bar{x} and \bar{y} are the means of characteristics of criminal and suspect separately, \bar{x} is the mean of characteristics of reference background population, σ is the standard deviation of the characteristics of criminal and suspect, a is the standard deviation of characteristic of reference background population, $z = (\bar{x} + \bar{y}) / 2$, $w = (m\bar{x} + n\bar{y}) / (m + n)$, m is the number of the criminal's characteristics observed, n is the number of the suspect's characteristics observed, $a = \sqrt{(1/m + 1/n)}$. In Lindley's formula, the second item represents the similarity between characteristic, and the third item represents the typicality of characteristics, and the first item represents how much larger the standard deviation of reference samples is compared to the standard deviation of the testing data.

The more similar the two samples are, the more likely they have come from the same source and the higher the ratio value will be. However, this must be balanced by samples' typicality. The more typical the two samples are, the more likely they have been taken randomly from the whole population, and the lower the ratio value will be. The value of likelihood ratio is therefore a result of interaction between the similarity and the typicality. Bayes' theorem makes it clear that both of these two aspects are necessary in evaluating the evidence. It is a very common fallacy to ignore typicality and assumes that the use of similarity alone is enough, which are not correct operation.

In forensic voice evidence, since voice features are mostly multidimensional, it is possible, in theory, to calculate likelihood ratio for each dimension and then combine them into an overall LR. The easy combination of LRs is one of the advantages of Bayes approach. The

combined LRs is the product of the each LRs based on the assumption that the features are independent. The approach is so-called "independence" or "Idiot's Bayes" LRs approach, which ignores correlation between variables.

The multivariate kernel-density method [13] can be used to calculate LR if the feature is multidimensional, but the theory of this method is too complex. It gradually was replaced by Gaussian mixture model (GMM), therefore, so this article only introduce the likelihood ratio calculation method based on GMM.

LR Calculation Method Based on GMM

Gaussian mixture model [14] approximates arbitrary probability distribution by a number of linear combination of the Gaussian probability density function, so it can be used to approximate the various distribution of phonetic characteristics.

The GMM can be regarded as a hybrid between a parametric density model and non parametric density model. Like a parametric model it has parameters that control the behavior of the density in known ways, but have no constraint that the data must be of a specific distribution type, such as Gaussian or Poisson distribution. Meanwhile, like a nonparametric model, the GMM has many degrees of freedom allowing arbitrary density modeling, without excessive computation and storage demands.

The advantages of applying a GMM as the likelihood function are that it is computationally inexpensive, and is based on a well-understood statistical model. For text-independent tasks, it is insensitive to the temporal aspects of the feature information, modeling only the underlying distribution of observations. GMM is the mainstream of statistical modeling, and has been widely used in automatic speaker recognition based on cepstral features vector system, and has good recognition performance. GMM and likelihood ratio are both based on statistical analysis, therefore, GMM has a natural inner link with LR, and very suitable for application in the calculation of the LR.

The following, taking speech evidence for example, clarify like li hood ratio calculation method based on GMM. The GMM used in acoustic-phonetics features models on the same voice unit, such as measuring the monophthong /a /. We should select dozens of units /a /, then mark their stability period and extract the features on the above marked stable periods. So, statistical

modeling on the same voice unit can reflect the variability of the marked units, and eliminate the influence of different phonetic units. Statistical model reflects the within source variability of selected features if you use the features from the same speaker. If you use the features from background speakers, statistical model reflects the between source variability. Then the value of likelihood is calculated on the above statistical models using questioned voice features, its ratio is the likelihood ratio. In calculating LR, suspect voice samples and background voice sample database are also needed in addition to the questioned voice samples. Calculation flow chart shown in figure 1.

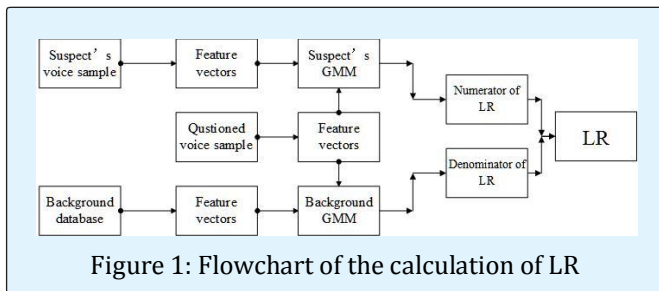


Figure 1: Flowchart of the calculation of LR

The extracted features of voice are generally multidimensional, and then the trained model is a multidimensional GMM. Because the quantity of questioned voice is usually not sufficient, the suspect's sample are often used to assess the within source variability. Evidence is not contained in the signal of speech, but is included in the degree of similarity between features extracted from the questioned voice and the suspect's voice. The degree of similarity is represented using likelihood value when feature vectors are compared with the GMM of another feature vectors [15].

These feature vectors of questioned voice are then used to compute the likelihoods on the above statistical models. The ratio of the two likelihoods is the LR. Mathematically, within source hypothesis is represented by parameter λ_{hyp} , which including the mean vector and covariance matrix parameters of the Gaussian distribution. The alternative between source hypothesis is

also represented by the GMM with the parameter λ_{hyp} . LR can be represented by formula (2).

$$p(X | \lambda_{hyp}) / p(X | \lambda_{hyp})$$

The numerator of LR quantifies the degree of similarity between the criminal and suspect samples, and the denominator of LR quantifies the degree of typicality of

the offender and suspect samples in the relevant population.

Conclusion

This paper introduced the necessity of quantitative evidence evaluation and the requirements of scientific evidence. We elaborated on the calculation methods and steps of quantitative analysis based on voice evidence, which proves that the quantification of evidence strength besides DNA is practicable and this method can also be extended to other disciplines of forensic science. Most important of all, the method is principle transparent, logical right way to make evidence evaluation scientific. It is easy understood and accepted by judges, juries, layers, fact-finders etc.

Up to now, many courts of some countries have adopted such examination opinions expressed in LR. In the criminal procedure law of China, examination conclusion has been modified to examination opinion, which gives an explicit direction that examination results will be reassessment in a more scientific way in the future. Therefore, evidence quantification has started and will be extended to other disciplines of forensic science.

References

1. Pillay KKS, Jester WA, Fox HA (1973) New Methods in Collection and Analysis of Gunshot Residues as a Forensic Evidence. J Abstr Pap Am Chem S (26): 23-23.
2. Porter J, Fouweather C (1975) Appraisal of Human Head Hair as Forensic Evidence. J SocCosmetChem 26(6): 299-313.
3. Downton P (1986) Forensic Evidence. J New Sci 112(1533): 68-68.
4. Peterson JL (1986) Factors Influencing the Adjudication of Felony Cases - What Role for Forensic Evidence. Abstr Pap Am Chem S 192: 2.
5. Savolainen P, Lundeberg J (1999) Forensic evidence based on mtDNA from dog and wolf hairs[J]. Journal of Forensic Sciences 44(1): 77-81.
6. Turner B, Wiltshire P (1999) Experimental validation of forensic evidence: a study of the decomposition of buried pigs in a heavy clay soil[J]. Forensic science international 101(2): 113-122.

7. U.S. Supreme Court. Daubert V. Merrel (1993) Dow Pharmaceuticals.
8. Anil_Alexander (2005) Forensic automatic speaker recognition. Ph.D, Indian Institute of Technology, Madras.
9. Saks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science [J]. Science, 309(5736): 892-895.
10. Wang Huapeng, Yang Jun, (2014) Automatic Speaker Recognition for Courtroom Based on Adaptive Within-Source-Variance Control. Journal of applied sciences 32(6): 582-587.
11. Morrison GS, Zhang C, Rose P (2011) An empirical estimate of the precision of likelihood ratios from a forensic-voice-comparison system. Forensic science international 208(1-3): 59-65.
12. Lindley D (1977) A problem in forensic science. Biometrika 64(2): 207-213.
13. Wang Huapeng (2013) Forensic Speaker Recognition. Doctor thesis, University of Chinese Academy of Sciences.
14. Reynolds DA, Rose RC (1995) Robust text-independent speaker identification using Gaussian mixture speaker models. Speech and Audio Processing. IEEE Transactions on 3(1): 72-83.
15. Morrison GS (2011) A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM). Speech Communication 53(2): 242-256.