International Research Journal of Computer Science (IRJCS)
Issue 08, Volume 4 (August 2017)
ISSN: 2393-9842
www.irjcs.com

THE EFFECT OF THE ANALYTICAL THRESHOLD ON THE LOSS AND GAIN OF DATA FOR SINGLE SOURCE AND MIXED DNA SAMPLES

Aidan E. Kirby, Jennifer Chance-Johnson and Heather Miller Coyle*
Forensic Science Department, Henry C. Lee College of Criminal Justice & Forensic Sciences,
University of New Haven, United States
Hcoyle@newhaven.edu

Manuscript History
Number: **IRJCS/RS/Vol.04/Issue08/AUCS10082**
DOI: 10.26562/IRJCS.2017.AUCS10082
Received: 09, August 2017
Final Correction: 14, August 2017
Final Accepted: 24, August 2017
Published: August 2017
Citation: E. Aidan, C. Jennifer, and H. Coyle. Research report, IRJCS, Issue VIII. Forensic Science Department, Henry C. Lee College of Criminal Justice & Forensic Sciences, University of New Haven, United States, Forensic Science Department, Henry C. Lee College of Criminal Justice & Forensic Sciences, University of New Haven, United States, *(*August 2017*)*
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India
Copyright: **©2017 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited**

Abstract—The forensic science community has been heavily scrutinized in recent years due to the release of the PCAST report. One of the methods under scrutiny is the analysis of DNA mixtures. These methods have been reassessed due to the lack of understanding, validation and methodical testing of the effects of changing the analytical threshold (AT) and the increasing use of probabilistic genotyping software programs. Cybergenetics states that their program, TrueAllele, is the most accurate statistically for DNA mixtures due to the lack of an analytical threshold in their DNA mixture assessment. By using no analytical threshold, the maximum number of peaks will be detected in a sample without data loss. Other software analysis programs allow for forensic laboratory policies that use differing analytical thresholds (e.g. 50 RFU, 75 RFU, 100 RFU) to screen for maximum confidence in allele calls compared to instrument noise. This study attempts to test the effects that the analytical threshold has on the percent loss of peaks for both single source DNA samples and DNA mixtures. Using the program GeneMarker where the analytical threshold can be manipulated, 103 mixed source and 53 single source DNA samples taken anonymously were evaluated with various analytical thresholds and the numbers of DNA fragments/peaks were tallied. For each data set, the only adjusted parameter was the analytical threshold. On average, single source samples had a 0.14% loss per threshold shift suggesting that there is little to no effect of the analytical threshold on the final number of allele calls. Conversely, mixed source DNA samples had almost twenty times that value, with a loss of data averaging 2.82%. The analytical threshold affects mixed source samples significantly more than single source samples and depending on the amount of total peaks in the DNA mixture and data quality (peak heights), there could be a substantial loss of data when analytical thresholds are used. Touch DNA samples are common types of forensic mixtures where there may be a major component but the minor contributors are not distinguishable. This could be due to data loss below the analytical threshold, reducing the number of informative loci for forensic reporting. The effect of this is to reduce clarity for number of contributors to a mixed sample and a loss in discrimination power or ability to identify the minor components in the mixture.

Keywords— forensic science, analytical threshold, allele drop-out, single-source DNA sample, DNA mixtures, DNA methylation

© 2014- 17, IRJCS- All Rights Reserved

I. INTRODUCTION

Due to the recent release of the PCAST report many of the forensic science community's methods have come under scrutiny [1]. One such method is the analysis of DNA mixtures [2-16]. Looking specifically at analytical thresholds (AT), there have been recent studies to attempt to clarify the effect of varying the AT on data [17, 18]. One published study by Bregu et al. looked at six methods for determining the optimal analytical threshold for a set of data. The results of this research stated that "[analytical thresholds] should be validated for each post-PCR procedure employed" and that "[analytical thresholds] derived from negatives should only be applied to samples with low levels of DNA" [17]. In another research study completed by Rakay et al., the effect of the analytical threshold on the rate of allele drop out was evaluated, and found that "[analytical thresholds] derived from baseline analysis of negatives were shown to decrease the frequency of allele drop-out by a factor of 100 without significantly increasing rates of erroneous noise detection" [18]. Cybergenetics has made claims in recent years that due to their program's lack of analytical threshold, all of the DNA data is used and, therefore, it is more accurate when compared to other probabilistic genotyping software programs [19].

There have been twelve admissibility hearings where courts have decided to allow TrueAllele DNA analysis (Cybergenetics Inc; Pittsburgh, PA) to be admitted in court and there will no doubt be more in the future (www.cybgen.com). The research presented here looks at the percent of alleles lost when the analytical threshold is raised for both mixed and single source samples. The research hypothesis was the analytical threshold would have (a) a greater effect on the mixed source samples than the single source samples and (b) there would be less than ten percent of data loss for each level of AT increase. This research used the software GeneMarker by SoftGenetics (State College, PA) to view and alter the analytical threshold parameters. The research hopes to further define the average percent of data loss when adjusting AT parameters during analysis. This is important since forensic science laboratories each set their own AT and the data would be analysed differently based on each of the forensic laboratory policies. In addition, although there is an obligation to reveal underlying data for full disclosure, often the data is not reanalysed electronically at different AT values unless by a special request of the laboratory personnel, by court order or by an independent expert.

II. MATERIALS AND METHODS

A collection of 54 single source samples and 103 mixed source samples were compiled. The single source samples were originally derived from anonymous buccal swabs collected from prior research projects conducted at the University of New Haven. The DNA mixtures were generated from prior touch DNA research to simulate crime scene samples. Each sample, mixed or single source, was analysed with GeneMarker software to view labelled peaks and edit the DNA profiles. The samples were analysed with the following settings: enhanced smooth, enhanced baseline subtraction, pull-up correction, spike removal and saturation repair. The DNA fragments were sized using the local southern sizing method. All samples were first tested with the analytical threshold at 0 relative fluorescence units (RFU) to visualize and assess all possible data collected during capillary electrophoresis. Samples were additionally tested with the threshold at 30 RFU (three times the common instrument baseline of 10 RFU) and at every addition of ten RFU until the threshold reached 70 RFU. At each AT value, the amount of alleles was recorded including each peak's allele, size and height. A percentage difference between the numbers of alleles detected at the higher AT compared to the numbers of alleles detected at the lower AT was calculated as percent data loss. A control DNA recovery study was performed for baseline data by extracting total human DNA with the QIA amp DNA Investigator kit (www.qiagen.com) using the manufacturer's recommended protocols for buccal swab, nail and hair samples. DNA quantitation was performed using the Quantifiler kit (www.thermofisher.com) per manufacturer protocol. The extraction negative control for the kit consisted of Tris-EDTA buffer without DNA and produced no detectable DNA; the extraction positive control consisted of buccal epithelial cells and recovered 20.7ng/ul of DNA. A methylation study was performed on buccal epithelial samples using a CpG MethylQuest DNA isolation kit (www.emdmillipore.com) to establish if there were gender differences in DNA recovery rate. Recovery rate is linked to AT usage, as the more DNA recovered, the greater the peak height in the final electropherogram. Touch DNA samples were collected using a double swab technique and processed and detected using the same methods as described for the control study for buccal swabs.

III. RESULTS

For the 53 single source samples, the average for the threshold difference was consistently below 1% data loss, with the average being a loss of 0.14% (Table 1). This small percentage difference illustrates that the AT has a minimal effect on single source buccal samples and very little if any "drop out" or data loss occurs. This is expected as buccal swabs are typically high quality samples with abundant DNA template for the polymerase chain reaction (PCR) process. Conversely, the mixed source samples stayed consistently around 2.5% loss, with the average being 2.82% (Table 1). The only threshold change within the mixed source samples to not have an average within the same range of the others is the 0-30 RFU shift.

© 2014- 17, IRJCS- All Rights Reserved

This is more than likely due to the fact that all the other AT intervals change by ten RFU, whereas the 0-30 RFU shift changes by thirty. If it was divided to contain 0-10, 10-20 and 20-30 RFU, respectively, there would more than likely be a similar pattern as with the 10 RFU interval comparisons. While a 2.82% loss is not an overly large percentage, it is significantly more loss than is seen in the single source samples. Depending on the size and complexity of the sample that could mean the loss of a multitude of peaks and with minor contributors, the loss can be significant. The larger difference between the threshold at zero and the threshold at thirty shows that if a sample was tested with a forensic laboratory AT of 30 RFU, the sample on average would be missing about 4% of the peaks close to the baseline. The AT effect is an average as the data loss is due to many variables such as number of contributors, level of DNA degradation, allele sharing and peak morphology due to data quality. In every case for DNA mixtures, the actual data loss is sample dependent but to obtain an estimate, this study has calculated an average loss of 2.5% for each interval of 10 RFU above the 30 RFU setting. With complex mixtures with numerous minor contributors and higher AT settings, the data loss can be as high as an estimated 18% in some casework simulated samples or as low as 0% with few minor components.

TABLE I. AVERAGE DATA LOSS PER AT INTERVAL

Threshold Changes (RFU)	Average Data Loss for Single Sources	Average Data Loss for Mixed Sources
0 – 30	-0.07%	-4.19%
30 - 40	0	-2.62%
40 - 50	-0.15%	-2.22%
50 - 60	-0.27%	-2.43%
60 – 70	-0.23%	-2.64%
Overall Average Data Loss:	-0.14%	-2.82%

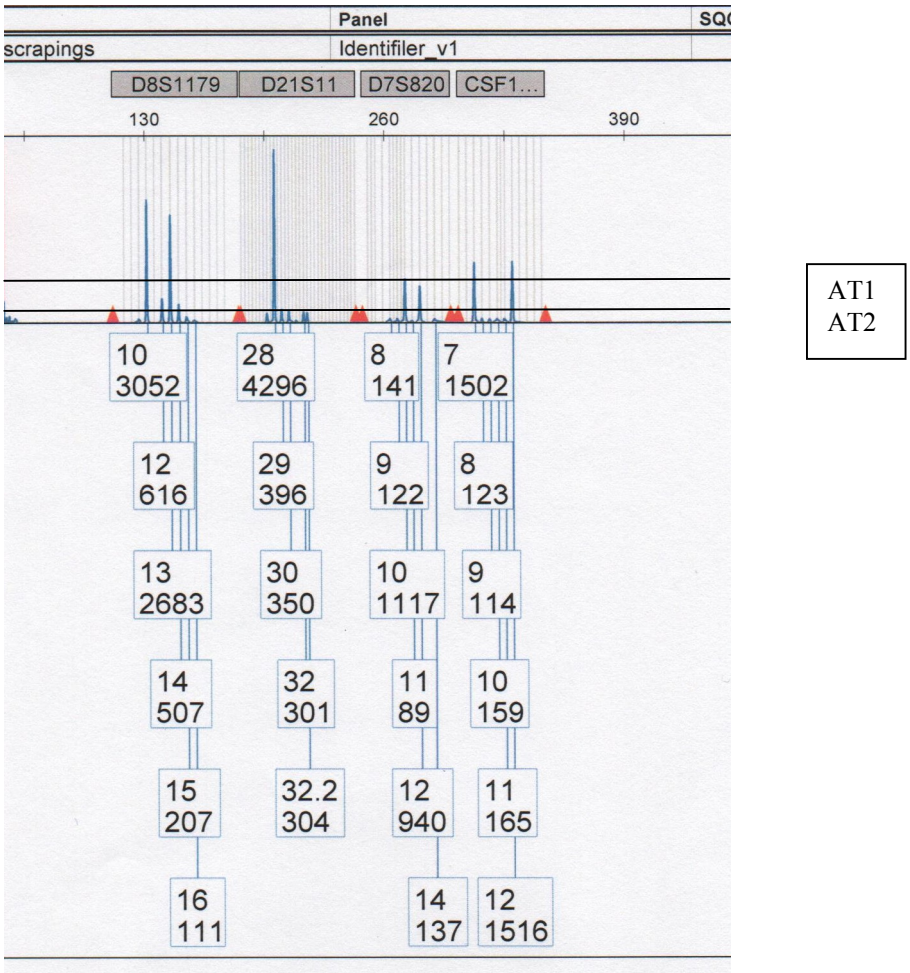

Fig. 1 DNA electropherogram depicting a complex DNA mixture with at least four donors for genetic loci D8S1179, D21S11, D7S820, and CSF1PO. If the analytical threshold (AT) were set at 100 RFU, then the 11 allele at the D7S820 locus would be lost data as the peak height (RFU) is 89 RFU. AT1 and AT2 illustrate how differing numbers of peaks would be labelled depending on where you set the AT level.

© 2014- 17, IRJCS- All Rights Reserved

Each peak or DNA fragment is labelled with the allele value (top) and the peak height (bottom) on the boxed label. This DNA profile is from hat scrapings where multiple individuals have either worn or handled the hat. This example illustrates the value in reviewing electronic data on a case by case basis for evaluation of low level peaks between 0-100 RFU that are clearly visible close to the baseline. For reference, the average DNA recovery from a buccal swab is 5.48 ng/ul, from liquid saliva is 3.40 ng/ul, and from fingernail clippings is 0.335 ng/ul (Table 2). Gender has no effect on the recovery rate as 98% and 96% methylated DNA was recovered from male and female buccal swabs (4 each), respectively. Recovery rates for DNA of high quality known reference samples and gender methylation patterns were evaluated as controls for comparison to the simulated crime scene touch DNA samples as proof of concept studies to illustrate kit functionality.

TABLE II. AVERAGE DNA RECOVERY

Sample Type (Single Source)	No. of Samples	Ave. Yield(ng/ul)	Range of Yield (ng/ul)	Collection Method
Buccal Swab	20	5.48	1.24 - 14.78	single cotton swab
Liquid Saliva	10	3.40	0.33 – 9.96	simulated swab
Fingernail Clippings	4	0.335	0.017 – 0.585	10 – 20 mg
Touch DNA	5	0.205	0.007-0.509	cotton swab

IV. DISCUSSION

Touch DNA is one form of forensic evidence that frequently results in complex DNA mixtures containing multiple contributors and minor donors. Touch DNA evidence is prevalent in property crimes and homicides. When analyzing and reviewing data, frequently peaks that could be authentic are visible in the electronic data below the AT. Each forensic laboratory sets their own policy for the AT to be used to analyze the data; therefore, there can be different answers as to the number of contributors by adjusting the AT as a variable [20-29]. Although the presence of additional peaks under the AT must be disclosed as being present, they may not be "reportable" (no allele calls) per the individual laboratory policy. This can be problematic with the use of probabilistic genotyping software [21, 25, 27, 30, 31] as many require an analyst's assessment as to the number of true contributors to calculate an accurate statistical match and the number may differ between the reportable above AT peaks and the unreportable below AT peaks based on individual laboratory policy. This concept is illustrated in Figure 1. Epithelial or touch DNA evidence can be defined as evidence with no visible staining that would likely contain DNA resulting from the transfer of epithelial cells from the skin to the object. Touch DNA is useful evidence to obtain from a crime scene, but entails limitations including the inability to test significantly small quantities of DNA with high levels of drop-out (missing data) and samples that may have four or more contributors (complex mixtures) from commonly handled items such as door knobs, computer keyboards, cell phones, and steering wheels [15, 19, 22]. Many factors related to the deposition of DNA may include washing of hands, amount of force applied, duration of contact with surface, angle of deposition, and genetics of the individual for cell sloughing (shedder status), the type of substrate, presence of perspiration and the biological evidence collection technique [32]. In addition to the direct deposition of touch DNA, DNA from other unknown sources on the hand can be transferred to another individual or object by secondary transfer. Secondary transfer can result in DNA transfer via one of three routes: person to person to object, person to object to object or person to object to person. Secondary transfer of other individual's DNA onto objects is another significant factor in creating complex DNA mixtures with multiple minor contributors that create challenges to accurately estimate the true number of contributors. This can be referred to as contamination, unaccounted for alleles, unidentified donors, or "drop-in"; added alleles to a sample that increase the complexity of the DNA mixture with the addition of donors via transfer events.

V. CONCLUSIONS

Overall, this study indicates that lowering the AT for data analysis of mixed source DNA samples such as touch DNA gains the analyst on average 2.82% of peaks, which depending on the amount of overall peaks per sample could be a substantial amount of information. As seen with an average of 0.14% data loss for single source samples per increase of 10 RFU AT intervals, the analytical threshold has minimal effect on high quantity DNA samples such as buccal swabs. This study suggests that the TrueAllele software claim of being more accurate due to a lack of analytical threshold appears to have merit, but only when in reference to mixed source DNA samples. This study also illustrates the importance of data disclosure to re-evaluate the electronic data for unlabelled peaks below the AT for potential additional contributors, for increased accuracy in statistics, for case reconstruction and for exculpatory data.

ACKNOWLEDGMENT

Thank you to the University of New Haven for funding this research and for providing the SoftGenetics GeneMarker software for the study. Thank you to Mesha Hazell-Smithen, Melissa Mellen and Stephanie Smith (University of New Haven) for their contributions with the DNA recovery and methylation studies.

© 2014- 17, IRJCS- All Rights Reserved

REFERENCES

1. Executive Office of the President, President's Council of Advisors on Science and Technology, Report to the President. "Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods," pp. 1-183, 2016.
2. C.H. Brenner, "Fairness in evaluating DNA mixtures," Forensic Sci. Int. Genet., vol. 27, p. 186, 2017.
3. K. Slooten, "Accurate assessment of the weight of evidence for DNA mixtures by integrating the likelihood ratio," Forensic Sci. Int. Genet., vol. 27, pp. 1-16, 2017.
4. F.R. Bieber, J.S. Buckleton, B. Budowle, J.M. Butler and M.D. Coble, "Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion," BMC Genet., vol. 17(1), pp. 125-140, 2016.
5. K/J. Slooten and T. Egeland, "Exclusion probabilities and likelihood ratios with applications to mixtures," Int. J. Legal Med., vol. 130(1): pp. 39-57, 2016.
6. P. Gill, H. Haned, O. Bleka, O. Hansson, G. Dørum and T. Egeland, "Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches-twenty years of research and development," Forensic Sci. Int. Genet., vol. 18, pp. 100-117, 2015.
7. K. Slooten, "Identifying common donors in DNA mixtures, with applications to database searches," Forensic Sci. Int. Genet., vol. 26, pp. 40-47, 2017.
8. C.D. Marsden, N. Rudin, K. Inman and K.E. Lohmueller, "An assessment of the information content of likelihood ratios derived from complex mixtures," Forensic Sci. Int. Genet., vol. 22, pp. 64-72, 2016.
9. G. Dørum, D. Kling, A. Tillmar, M.D. Vigeland and T. Egeland, "Mixtures with relatives and linked markers," Int. J. Legal Med., vol. 130(3), pp. 621-634, 2016.
10. M.W. Perlin, "Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information," J. Pathol. Inform., vol. 6, p. 59-69, 2015.
11. K. Slooten, "Distinguishing between donors and their relatives in complex DNA mixtures with binary models," Forensic Sci. Int. Genet., vol. 21, pp. 95-109, 2016.
12. D. Taylor, J.A. Bright and J. Buckleton, "Considering relatives when assessing the evidential strength of mixed DNA profiles," Forensic Sci. Int. Genet., vol. 13, pp. 259-263, 2014.
13. D. Taylor, J.A. Bright and J. Buckleton, "The 'factor of two' issue in mixed DNA profiles," J. Theor. Biol., vol. 363, pp. 300-306, 2014.
14. D. Taylor, J. Buckleton and I. Evett, "Testing likelihood ratios produced from complex DNA profiles, " Forensic Sci. Int. Genet., vol. 16, pp. 165-171, 2015.
15. H. Haned, C.C. Benschop, P.D. Gill and T. Sijen, "Complex DNA mixture analysis in a forensic context: evaluating the probative value using a likelihood ratio model," Forensic Sci. Int. Genet., vol. 16, pp. 17-25, 2015.
16. J. Isaacson, E. Schwoebel, A. Shcherbina, D. Ricke, J. Harper, M. Petrovick, J. Bobrow, T. Boettcher, B. Helfer, C. Zook and E. Wack, "Robust detection of individual forensic profiles in DNA mixtures," Forensic Sci. Int. Genet., vol. 14, pp. 31-37, 2015.
17. J. Bregu, D. Conklin, E. Coronado, M. Terrill, R.W. Cotton and C. M. Grgicak, "Analytical thresholds and sensitivity: establishing RFU thresholds for forensic DNA analysis, J. Forensic Sci., vol. 58, 120–129, 2012.
18. C.A. Rakay, J. Bregu and C. M. Grgicak, "Maximizing allele detection: effects of analytical threshold and DNA levels on rates of allele and locus drop-out," Forensic Sci. Int. Genet., vol. 6, pp. 723–728, 2012.
19. M.W. Perlin, J.M. Hornyak, G. Sugimoto and K.W. Miller, "TrueAllele(®) genotype identification on DNA mixtures containing up to five unknown contributors," J. Forensic Sci., vol. 60(4), pp. 857-868, 2015.
20. T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J.S. Buckleton, J.A. Bright, D.A.Taylor and A.J. Onorato, "Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles," Forensic Sci. Int. Genet., vol. 29, pp. 126-144, 2017.
21. M.A. Marciano and J.D. Adelman, "PACE: probabilistic assessment for contributor estimation- a machine learning-based assessment of the number of contributors in DNA mixtures," Forensic Sci. Int. Genet., vol. 27, pp.82-91, 2017.
22. L.E. Alfonse, G. Tejada, H. Swaminathan, D.S. Lun and C.M. Grgicak, "Inferring the number of contributors to complex DNA mixtures using three methods: exploring the limits of low-template DNA interpretation," J. Forensic Sci., vol. 62(2), pp. 308-316, 2017.
23. M.D. Coble, J. Buckleton, J.M. Butler, T. Egeland, R. Fimmers, P. Gill, L. Gusmão, B. Guttman, M. Krawczak, N. Morling, W. Parson, N. Pinto, P.M. Schneider, S.T. Sherry, S. Willuweit and M. Prinz, "DNA Commission of the International Society for Forensic Genetics: recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications," Forensic Sci. Int. Genet., vol. 25, pp. 191-197, 2016.

© 2014- 17, IRJCS- All Rights Reserved

24.Ø. Bleka, C.C. Benschop, G. Storvik, and P. Gill, "A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles," Forensic Sci. Int. Genet., vol. 25, pp. 85-96, 2016.
25.H. Swaminathan, A. Garg, C.M. Grgicak, M. Medard, and D.S. Lun, "CEESIt: A computational tool for the interpretation of STR mixtures," Forensic Sci. Int. Genet., vol. 22, pp. 149-160, 2016.
26.C.C. Benschop, H. Haned, L. Jeurissen, P.D. Gill and T. Sijen, "The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures," Forensic Sci. Int. Genet., vol. 19, pp. 92-99, 2015.
27.H. Swaminathan, C.M. Grgicak, M. Medard, and DS Lun, "NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping," Forensic Sci. Int. Genet., vol. 16, pp. 172-180, 2015.
28.I.W. Evett and S. Pope, "Is it to the advantage of a defendant to infer a greater number of contributors to a questioned sample than is necessary to explain the observed DNA profile?," Sci. Justice, vol. 54(5), pp. 373-374, 2014.
29.D. Taylor, J. A. Bright and J. Buckleton, "Interpreting forensic DNA profiling evidence without specifying the number of contributors," Forensic Sci. Int. Genet., vol. 13, pp. 269-280, 2014.
30.H. Miller Coyle, "Sources of computational error in probabilistic genotyping software used for DNA mixture interpretation," International Research Journal of Computer Science, vol. 2(5), pp. 12-16, 2015.
31.K. B. Watters and H. Miller Coyle, "Forensic statistical tool (FST): a probabilistic genotyping software program for human identification," Jurimetrics, vol. 56 (2), pp. 183-195, 2016.
32.M. Hazell-Smithen, T. Callahan, and H. Miller Coyle, "Touch DNA and the ability to detect the correct source," International Journal of Advanced Technology and Science, vol. 1 (1), pp. 45-51, 2014.

© 2014- 17, IRJCS- All Rights Reserved