C-BET evaluation of voice biometrics

Dmitry O. Gorodnichy*, Michael Thieme,
David Bissessar, Jessica Chung, Elan Dubrofsky, Jonathon Lee

***Abstract*— C-BET is the Comprehensive Biometrics Evaluation Toolkit developed by CBSA in order to analyze the suitability of biometric systems for fully-automated border/access control applications. Following the multi-order score analysis and the threshold-validated analysis defined within the C-BET framework, the paper presents the results of the C-BET evaluation of a commercial voice biometric product. In addition to error tradeoff and ranking curves traditionally reported elsewhere, the paper presents the results on the newly introduced performance metrics: threshold-validated recognition ranking and non-confident decisions due to multiple threshold-validated scores. The results are obtained on over a million voice audio clip comparisons. Good biometric evaluation practices offered within C-BET framework are presented.**

Disclaimer: Specific hardware and software products identified in this paper were used in order to perform the evaluations described in this document. In no way does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the Canada Border Services Agency, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

1. INTRODUCTION

With thousands of people crossing country borders every minute and with the ever increasing need to make cross-border travel both secure and efficient, biometric-enabled Automated Border Control (ABC) is seen as one of the most promising applications of biometrics [1,2].

Currently, the main modalities used at the border are fingerprints, iris, and face, due to their high performance levels and traditional border/passport control practices. The use of these modalities requires active "stand-by" participation of the travelers. At the same time as discussed in [3], it is anticipated that border control applications may also require or may benefit from the use of biometrics in "stand-off" mode, in which a biometric sample is captured without a person's participation and possibly even without a person's awareness of the fact that his/her biometric sample is captured.

The benefit of "stand-off" biometrics is seen in its ability to facilitate instant identification of individuals without their direct engagement, which could improve the overall positive experience of bona fide travelers using the system, and which could also be used to identify travelers with a criminal record.

Voice is one of the most accessible biometric modality that can be easily captured and analyzed while a traveler is presenting himself to the border officer. Voice has the potential to improve or validate the recognition results obtained with other biometrics, and it may also provide critical means for validating a person's identity in unmanned points of entry, provided that its performance is trusted.

Dmitry O. Gorodnichy, Dave Bissessar, Elan Dubrofsky, Jonathon Lee with the Video Surveillance and Biometric (VSB) section of the Science and Engineering Directorate of the Canada Border Services Agency (CBSA-S&E), 14 Colonnade, Ottawa, ON, K2E 6T7, Canada. Michael Thieme and Jessica Chung are with International Biometric Group (IBG), 1 Battery Park Plaza, New York, NY 10004, USA. Corresponding author (Email: dmitry.gorodnichy@cbsa-asfc.gc.ca, Tel: (613) 954 3785, Fax: (613) 960-5184).

Citation: Dmitry O. Gorodnichy, Michael Thieme, Dave Bissessar, Jessica Chung, Elan Dubrofsky, Jonathon Lee . "C-BET evaluation of voice biometrics", SPIE Defence, Security & Sensing Conference, Special Track on Biometric Technology for Human Identification VIII (DS108), Orlando, April 25-29, 2011.

In this paper, we present the results of the comprehensive evaluation of the voice biometric modality conducted recently by CBSA and IBG with the objective to better understand the properties of this modality in terms of its recognition performance, its recognition decision confidences, and ultimately, its applicability to border applications. The evaluation was conducted following the multi-order score analysis and the threshold-validated analysis defined within the C-BET (comprehensive biometrics evaluation toolkit) framework developed by the CBSA [3-7]. This presentation therefore also serves the purpose of illustrating the concepts and the utility of the C-BET framework for a wider research and industry community, using a voice biometric product as an example.

The paper is organized as follows. Next section presents a brief summary of traditional voice biometric applications, conventional biometric evaluation methodologies, and the motivation for the development of the new comprehensive evaluation methodology. Good evaluation practices and a detailed description of the multi-order score analysis and the threshold-validated analysis defined within the C-BET framework are presented in Section 3. The application of the C-BET framework to the evaluation of a commercial start-of-art voice biometrics system is presented in Section 4, which also describes the voice data corpora used for this evaluation. The C-BET two-page summary report obtained for the voice modality is presented in the Appendix. Discussions on the implications of the obtained results conclude the paper.

2. TRADITIONAL VOICE BIOMETRICS EVALUATIONS AND MOTIVATION FOR BETTER EVALUATION METHODOLOGY

Voice biometrics has been used for several years in telephony by many banks and correction services - to verify the identity of calling individuals (one-to-one speaker verification). It is also used by police in forensic investigations as an important piece of evidence (one-to-many speaker identification) [9]. It has never been used however in fully-automated applications and in particular those dealing with high security.

Most referenced evaluations of voice biometrics, Speaker Recognition Evaluation (SRE), have been conducted over the years by NIST [10-13]. The methodology used in these evaluations is based on the traditional applications of the voice biometrics such as those listed above. The corpora of voice data is created and then used either i) to compute the verification match/non-match rates and the corresponding Detection Error Trade-off (DET) and Receiver Operator Characteristic (ROC) curves – for one-to-one speaker verification systems, or ii) to compute identification rates and the corresponding Cumulative Match Characteristic (CMC) curves – one-to-many speaker identification systems.

Such *application-targeted evaluation* methodology is common for all biometric modalities. - Following the evaluation methodology used by NIST [10-13], industry [14-16] and academia [17], one can notice that each evaluation is normally tailored for a particular application that the biometric system is applied to. If the system's intended use is to help a human analyst to recognize an individual, then the system is referred to as a *1-to-N identification system* and CMC curves are computed to evaluate the performance of this system. On the other hand, if the system is developed for access control, then it would be called a *1-to-1 verification* system and the evaluation of this system will be largely limited to computing DET/ROC curves only.

However, what if the application of the system is not known? What if a modality that has been conventionally used for forensic (non-automated) purpose only is now evaluated for its suitability in operation in fully-automated mode for access/border control, which is the case with the Face modality? Or when a constrained modality that has been traditionally used in fully-automated access/border control is now tested in unconstrained environments where fully automated recognition decision may not be possible, as in evaluation of Iris at Distance (on the move) biometric technology? Which Figures of Merit curves and statistics should be used then?

In response to these questions and driven by the operational need to better understand the technology that could be potentially deployed in the field, the Comprehensive Biometric Evaluation Toolkit (C-BET) evaluation framework has been proposed in [4-6] and further refined in [3]. The framework is designed to supplement the results that would normally be reported elsewhere (eg. NIST), but so that to provide a deeper understanding or "better feel" of the "Black Box", which the biometric system is, through the investigation of all higher detail information that could be possibly extracted and inducted through the experimentation with the system.

In the conception of the C-BET framework, another important observation related to the biometric evaluation process was instrumental. Large scale evaluation of a biometric system can be considered as a three-stage process. The amount of time and effort required to prepare the testing datasets that contain large number of Genuine and Impostor sample pairs (Stage 1) and the amount of time and effort required to learn a biometric product and to have it run on the entire dataset to obtain all Genuine and Impostor comparison scores (Stage 2) is normally much more significant than that of the final task of processing all computed scores and reporting the obtained performance statistics and graphs (Stage 3).

> *It appears therefore unfortunate that after an immense effort invested in the first two stages of the evaluation process leading to the computation of all scores, it is only a fraction of the statistical analysis, which could be potentially obtained from all computed scores, that is reported. In many cases, once an evaluation report is published, the score data that was used to generate the report will be discarded, and neither the user nor the developer of the system will ever know "the rest of the story"!*

The C-BET evaluation methodology and the toolkit, which is being made available from the DRDC portal [19] for registered users, are developed so that to allow one to report "the entire story" about a biometric system's performance.

3. MULTI-ORDER BIOMETRIC SCORE ANALYSIS

Multi-order score analysis is introduced in [4-6] and further refined in [3] as an important biometric performance methodology to allow the investigation into the risks and risk mitigation factors related to having non-confidence outputs in fully-automated biometrics systems. It was inspired by and originally applied to the evaluation of commercial iris biometrics systems such as those that can be potentially used for the CBSA-operated NEXUS traveler program [20].

The multi-order terminology for biometric score analysis comes from the analogy with multi-order statistics, in which Order-0 statistics signifies using the value itself, order-1 statistics signifies computing the average of several values, and Order-2 and Order-3 statistics signify computing the deviation (variance) and high-order statistical moments. Similar to statistics, the scores of a biometric system can be analyzed at several levels (or orders) of detail to provide incrementally more information for better decision making - in designing a system and its evaluation.

When used within a comprehensive performance evaluation procedure [3-6], the multi-order score analysis is shown to provide additional insights on a system's performance and reliability and expose the risks due to non-confident recognition decisions.

As shown in [3,7-8], it can also be used to improve the overall performance of a biometric system, when applied as a post-processing score recalibration filter. The usage of the multi-order score analysis terminology for biometric system design is illustrated by the following example.

Consider an iris recognition system with the matching threshold set at T = 0.33. When a probe iris image is compared against the images of five (different) people in an enrollment database, five matching scores are obtained (0.45, 0.32, 0.47, 0.34, 0.31). The Order-1 system, which takes the decision based on the assumption that being below a threshold is sufficient for the recognition decision, finds the first score below the threshold (0.32) and reports match for the 2nd person. The Order-2 system finds the smallest score (0.31) and reports the match for the 5th person. The Order-3 system however will not simply report the match but would also assign a confidence value to the match based on all score information available, which in this case will be low, since the score of the 4th person (0.34) is also close to the threshold.

In the following, we further summarize the key definitions and recommendations of the C-BET multi-order score analysis from [3] and present their application to the evaluation of voice modality using a commercial speaker recognition system.

3.1 Definitions, good evaluation practices and examples

First, ***Order-0 score analysis*** is defined as visual statistics and graphical visualization related to score distributions. Such visual analysis does not produce a performance metric in itself, yet it is found very useful to provide insights on how a system performs and where the performance bottlenecks could be.

Order-0 visual score analysis is shown in Figure A.3. Such visual analysis about an unknown ``Black Box" biometric system should always be obtained first, prior to further examination of the system, because it reveals the inner properties of the system. Particularly, it can be used to obtain the a-priori probabilistic distributions of Genuine and Impostor matching scores, which can then be used to maximize the probability of more reliable decisions using the Gorodnichy-Hoshino post-processing score calibration technique [7-8]. It also summarizes the properties of the dataset, such as the number of genuine and impostor comparisons used in the evaluation, which can be used to obtain the FMR/FNMR confidence bounds.

Definition: ***Order-1 score analysis*** *is based on computing and analyzing a single matching score, as in traditional 1-to-1 verification systems and when plotting DET/ROC curves.*

In traditional terminology, Order-1 analysis can be viewed and referred to as the ``verification analysis", which is conventionally performed for fully-automated access/border control systems.

Figures A.4-6 show Order-1 score analysis results. When plotting DET / ROC curves, all measured points should be explicitly shown on the curve. Showing only the extrapolated curves may mislead people into believing that certain rates are achievable by a system, when they are not. Showing the measured points can also serve to validate the appropriateness of the threshold increments used in conducting the evaluation.

Additionally, to avoid misinterpretation, for plots drawn using logarithmic scales, it is recommended to mark points corresponding to FMR / FNMR equal to zero as **Virtual 0**, as shown in Figure A.4.

Definition: ***Order-2 score analysis*** *is based on computing multiple matching scores and analyzing the score ranking statistics (or best K scores), as in 1-to- N comparisons used in investigative-mode recognition and when plotting the CMC curves.*

In traditional terminology, Order-2 analysis can be viewed and referred to as the ``identification analysis", which is conventionally performed for 1-to-N investigative systems.

Figures A.9-10 show Order-2 score analysis, which plots the number of instances when the genuine match was the best, second best etc. These curves can be seen as the derivative of the traditional CMC curves, which are used for evaluation of biometric identification systems for forensic purposes. These curves show the integral value of the identification rank, indicating that the genuine score was among the best K scores without specifying whether it was the K^{th}-best, $K\text{-}1^{th}$ best or the best score.

The reason for plotting Order-2 score curves as shown in Figures A.9-10 and not as traditional CMC is to tell more about the system. It also allows us to apply the innovative *Threshold-validated* terminology described below, according to which each matching result is labeled either as *Threshold-validated (TV)* or *non-Threshold-validated (non-TV)* depending on whether it passed a comparison to the threshold or not, i.e. whether it is smaller (or higher, depending on system design) than the matching threshold.

Definition: *The matching score of a biometric system is defined as* ***Threshold-validated*** *if it passes a comparison to the system matching threshold.*

Such definition is introduced to avoid referring to the scores as ``Matched" or ``Accepted" (as they are traditionally called in 1-to-1 verification systems), because the final ``Match" / ``Accept" system decision with the high-order score analysis may not only be based on the score comparison to the threshold, but other factors such as confidence.

In particular as described in [3], the concept of Threshold-validated biometric identification becomes vital when designing card-less / input-less biometric-enabled access and border control systems, in which a person's identity is recognized from a list of pre-approved enrolled individuals. It also becomes very important when evaluating the applicability of an identification biometric system that has been conventionally used for manual forensic examination only (such as in face recognition used by police) to new applications of it, such as automated triaging or tagging of faces in video, which are now becoming also of high interest for border agencies.

For an open dataset, when a probe sample is not in the dataset (which is the case in the evaluation presented in this paper) indicated as Rank 0 in Figures A.9-10, the Threshold-validated analysis provides the information on the likelihood that a random person can pass through the system. .

The Threshold-validated terminology becomes also very useful in defining and applying Order-3 score analysis, which looks at all relationship among the scores and their relationship with respect to the threshold.

Definition: ***Order-3 score analysis*** *is based on computing / analyzing the relationship between the match scores, as when finding the difference between the best and second-best match scores, finding all threshold-validated scores, or when applying the post-processing recalibration of the scores based on 1-to-N comparisons.*

The Order-3 score analysis results are shown in Figures A.2, 7-8, 11-12, 13-14. Figure A.2 shows the Performance Report Summary table, the third column of which indicates the Failure of Confidence Rate for each threshold, which according to its definition in [13] is the number of instances when there was more than one Threshold-Validated match for a probe.

Figure A.7-8 show the rate of *recognition confidence* measured in terms of the normalized distance from the best score to the second best score – for genuine best matches (left image) and for imposter best matches (right image). Ideally, one would like to have low confidence for best score if it belongs to an imposter, and high confidence if it belongs to a genuine comparison.

Figure A.11-12 shows statistics on the number of Threshold-validated matches. Additionally, the number of those Threshold-validated matches that scored the best are marked for each genuine and imposter match. Ideally for a fully automated system, one would like to have only one Threshold-validated match and this match to be the best and to correspond to the genuine comparison (Yellow colour). If it corresponds to the imposter comparison (Blue colour), then such system cannot be used for automated border/access control. At the same time, even if there is more than one Threshold-validated match, but the genuine score is the best (Blue colour), then such a system has a good potential to be used for automated decision making, provided, of course, that it is designed to maximize the confidence of its decisions through the use of Order-3 score analysis.

Finally, Figure A.13-14 shows the simplified representation of the Threshold-validated Order-3 analysis, in which the statistics with respect to the six possible performance outcomes is shown:

1. G (bt) T (bt) I | 2. G (bt) I (bt) T ‖ 3. I (bt) G (bt) T | 4. I (bt) T (bt) G ‖ 5. T (bt) G (bt) I | 6. T (bt) I (bt) G

where G signifies the Genuine score, I signifies Imposter score, T stands for Threshold, (bt) signifies “better than” in the biometric sense.

The six possible outcomes of the threshold-based statistics are sorted according to their meaning for the system performance description. - Outcome 1 is an ideal outcome for a fully-automated system. Outcome 4 is the worst outcome. Outcomes 2 and 3 indicate that additional processing is required for the system to be operational – either done by a human analyst, or done by a computer through the higher order score analysis. Outcomes 5 and 6 are indicative of the fact that either the data is not reliable or the biometric modality is not sufficiently constrained.

4. EXPERIMENTAL SETUP AND DATA

The voice biometrics evaluation was conducted using the Automatic Speaker Identification System (ASIS) manufactured by Agnitio [18]. ASIS is a text-, language-, and channel-independent speaker identification system designed to provide centralized identification services across large voice databases.

ASIS performs 1:N searches against a database of enrolled audio files, returning a rank list of candidates (as well as comparison scores) that most closely match a given subject.

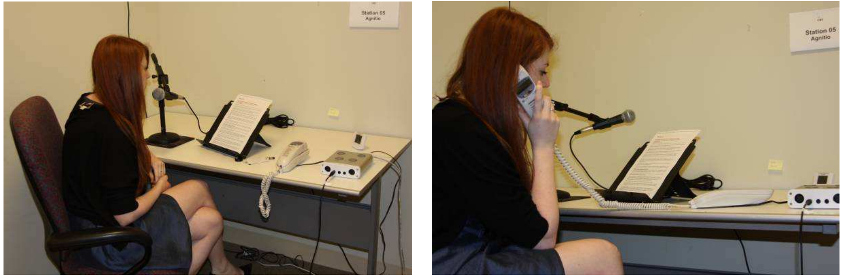


Fig. 1. The setup used to collect audio data corpora for voice biometric modality evaluation.

The Agnitio workstation was used to collect voice recordings through a microphone (using a Shure SM58 microphone[†]) and a telephone (using a Northwestern Bell NWB EasyTouch 77519 telephone[‡]), as shown in Figure 1.

Test Operators adjusted the height and orientation of the microphone to ensure that the Test Subject could comfortably read printed text. Test Operators were permitted to terminate and restart recordings if Test Subjects were reading too softly.

Two visits were arranged, with the total of 1832 audio recordings collected: 1010 - at the first visit, and 812 - at the

[†] http://www.shure.com/proaudio/products/wiredmicrophones/us_pro_sm58-cn_content

[‡] http://www.ahernstore.com/nwb-77519.html

second visit.

A 60-second probe and a 15-second extract was created for each enrollee at each visit. Thus in total four datasets were created– two for each audio channel: 15 mic, 60 mic, 15 tel, and 60 tel, in which a voice sample from each enrollee is present either once (if s/he participated in one visit only), or twice (if s/he participated in both visits).

Agnitio matching accuracy is rendered as ID Rate. Two probe durations were tested: the full 60-second probe and a 15-second extract for each audio source (mic and tel).

Figures 2 shows the CMC curves for each audio source for each of four possible audio comparision results: (tel vs. tel, tel vs. mic, mic vs. tel and mic vs. mic).

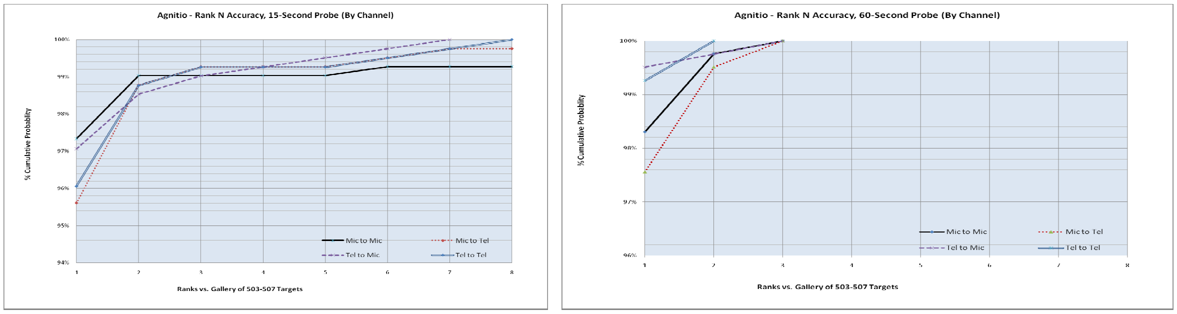


Fig. 2. CMC curves for 15 sec and 60 sec voice datasets .

4.1 Dataset used for C-BET evaluation

For more detailed analysis of the voice modality, a number of subsets were created from the set of all comparison scores to be analyzed using the C-BET methodology. For each sample duration (15 seconds and 60 seconds), a score subset was created for each possible comparison (tel vs. tel, tel vs. mic, mic vs. tel and mic vs. mic). Thus, we were able to perform the C-BET analysis on each subset of data in order to analyze each subset independently and compare them to each other.

Each score subset contains 65536 comparisons made up of two visit IDs and the associated comparison score[§]. The visit IDs are mapped to their associated subjects so that it can be determined whether both recordings are of the same person or not.

The results reported by C-BET evaluation for voice biometric modality using this dataset are presented in Appendix A. To evaluate the feasibility of using voice biometrics for fully automated decision-making the threshold-based statistics is computed using the simplified Threshold-validated Order-3 analysis described above. Figure 3 shows the aggregated threshold-based statistics computed on all voice comparisons (i.e. all tel vs. tel, tel vs. mic, mic vs. tel and mic vs. mic comparisons) for the complete audio dataset and for a subset of the dataset used for the C-BET evaluation.

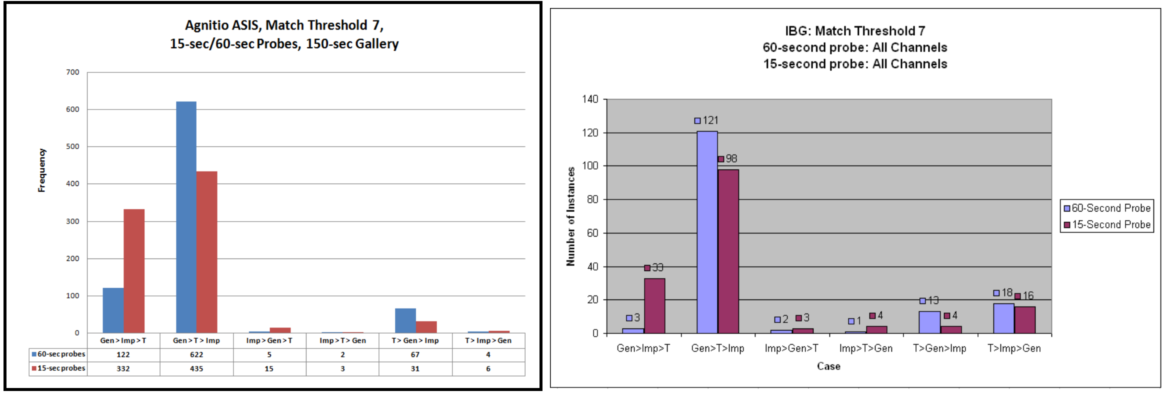


Fig. 3. Aggregated threshold-based statistics computed on the full dataset and for a subset of it used for the C-BET evaluation, the results of which are shown in Appendix A.

[§] The number of scores used for each subset (65536) is due to the limitation of MS Excel format which was used to transfer data files during the experiments.

5. DISCUSSION

5.1 C-BET implementation and the Visual Analytics side of it

Visual Analytics is a growing area of science and technology aimed at providing professionals with the efficient means to analyze large amounts of data.

When, as a result of a large-scale biometrics evaluation, millions of scores are obtained by comparing large numbers of genuine and imposter data, it becomes a big challenge for an analyst to analyze the data so as to make the best decision related to system selection and/or tuning. Traditionally these scores are used to generate detection error tradeoff and/or Cumulative Match Characteristic ranking curves, which are then used to visually compare systems to one. As discussed, these curves alone however do not provide complete knowledge about the system. Particularly, they do not facilitate the analysis of the risk of fully-automated border/access control biometrics systems due to non-confident matches, nor do they help to fine-tune the systems with respect to the factors that affect system robustness.

In order to provide a complete picture of the biometric system performance, the Comprehensive Biometrics Evaluation Toolkit (C-BET) developed by the Scientific and Engineering Directorate of the Canada Border Services Agency (CBSA-S&E) can be used.

To facilitate biometric system performance analysis and tuning using C-BET methodology, CBSA-S&E has develop a JAVA program software that allows one to automatically generate all C-BET metrics and graphs for a set of biometric scores. The C-BET software takes all scores obtained in a large scale evaluation to instantaneously generate multi-sheet MS EXCEL files containing easy to browse and analyze graphs related to the system performance.

As such, the C-BET software allows one to efficiently compare biometrics systems to one another and quickly investigate the affect of different parameters, such as the match threshold, on the system performance, by visually comparing images to one another. It also allows one to choose the parameters that yield the optimal system performance, which occurs when both of the following conditions are met: a) the genuine match scores are the best (Rank-1) scores, and b) the genuine match scores are the *only* threshold-validated scores.

For the voice biometric system used in this study, Figures A.9-14 show how to select the best threshold by visually comparing multi-order analysis graphs to one another. Particularly, it can been seen the performance of the system in the right column of the page, which corresponds to the threshold equal to 5.0, is better than that of the system in the left column of the page, which corresponds to the threshold equal to 4.0, in terms of the number of unreliable genuine matches and the number of non-confident imposter matches.

5.2 Implications of the obtained results and future prospects

In this paper, the C-BET evaluation framework is applied to examine the applicability of voice biometric modality for fully-automated border/access control applications. While the obtained results may not be considered conclusive, due to the limitations of the used audio data corpora, they provide a good indication on the limitations and potential of voice biometric systems.

The obtained threshold-validated multi-order score analysis results indicate that the performance of voice biometric system is comparable to those traditionally used for border/access control. The majority of cases the genuine match is the only threshold-validated match and has also best score. In cases, when there is more than one threshold-validated match, the system may still be applicable for border/access control, provided that the system is properly designed to deal with non-confident matches. This indicates that voice biometric systems may now be considered for further testing and piloting in fully-automated biometric recognition applications in real-environments, either as a single-modality device with additional post-processing filter or combined with other biometrics modalities such as face recognition.

Acknowledgement

The scores of voice biometric comparisons for this study have been provided to CBSA by IBG under CBSA-led project PSTP08-0110BIO “Biometric Border Security Evaluation Framework” funded by the Defence Research and Development Canada’s Center for Security Science (DRDC-CSS). The partial support of the DRDC-CSS for the development of C-BET is gratefully acknowledged.

APPENDIX A.

Appendix A shows a two-page summary report of the C-BET evaluation of a voice biometric system. Page 1 shows the basic system performance results based on the the Order-0 and Order-1 score analysis:

1) Description of the product and dataset used in the evaluation;
2) Performance Summary table showing most important evaluation metrics (FTA, FCR, and FNMR at given FMR);
3) Distributions of Genuine and Impostor matching scores obtained by the system (Order-0 analysis);
4) DET curves with and without calibration (Order-1 analysis);
5) FMR/FNMR distributions (Order-1 analysis),
6) FMR/FNMR distributions zoomed on the area of highest importance; and
7-8) Order-3 recognition confidence statistics -- for genuine best matches and for impostor best matches, where confidence is measured as the normalized distance from the best score to the second best score.

Page 2 shows the threshold-validated Order-2 and Order-3 score analysis -- for two different threshold values:

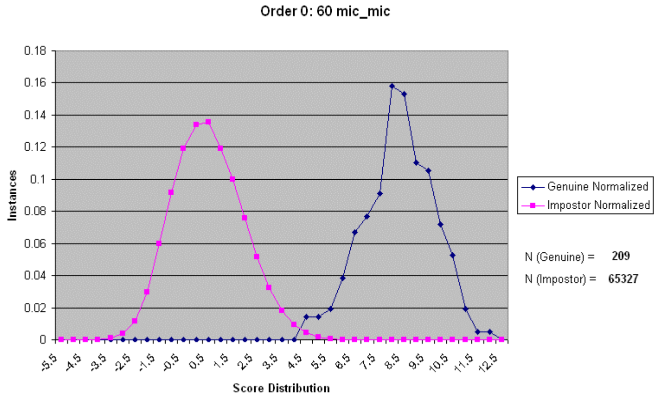
9-10) The number of instances when the genuine match was the best, second best etc.;
11-12) The statistics on the number of Threshold-validated matches and those of them that ranked the best;
13-14) Six-outcomes of the summarized threshold-based analysis.

For more information on the used metrics and C-BET evaluation methodology, see [3].

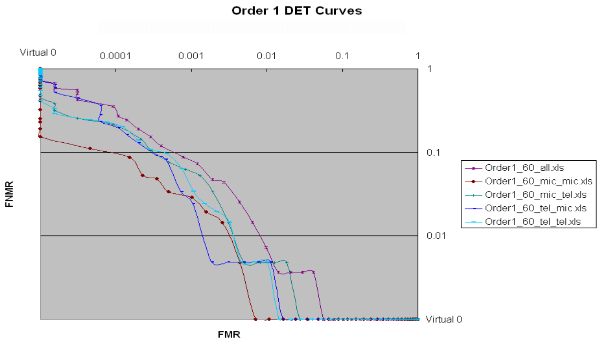
1)

Thresh	FMR	FNMR	FCR
4.25	**0.01072**	**0.00000**	**0.56369**
4.50	0.00710	0.00100	0.46815
4.75	0.00444	0.00478	0.35987
5.00	0.00260	0.01435	0.23885
5.25	0.00159	0.01914	0.17197
5.50	0.00103	0.02871	0.11783
5.75	0.00051	0.03349	0.06369
6.00	0.00035	0.04785	0.04459
6.25	0.00023	0.05263	0.02866
6.50	0.00015	0.08612	0.02548
6.75	0.00005	0.11005	0.00955
7.00	0.00001	0.15311	0.00000
7.25	**0.00000**	**0.19139**	**0.00000**

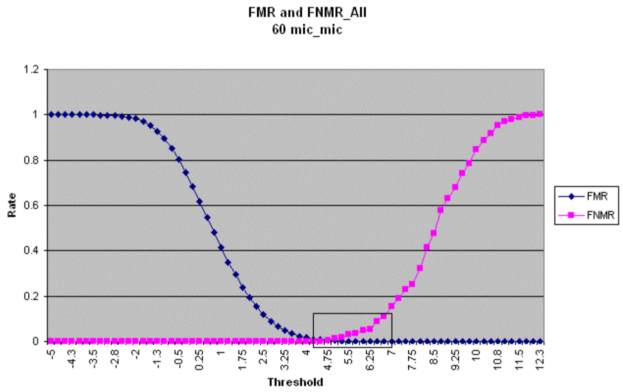
2)


Order 0: 60 mic_mic

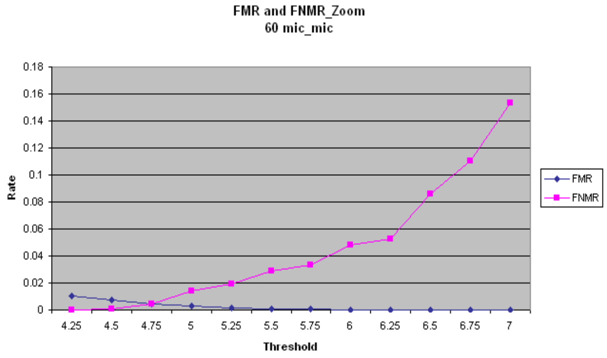
3)


Order 1 DET Curves

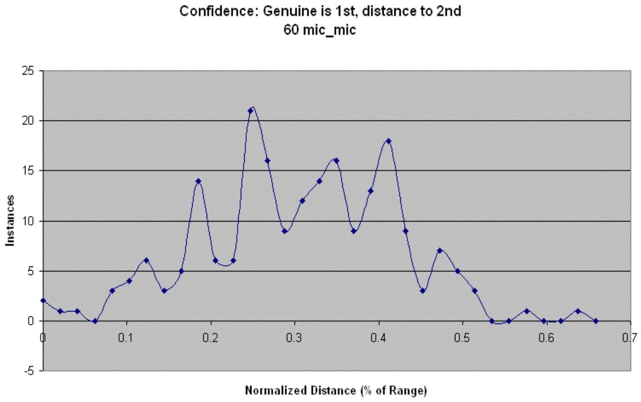
4)


FMR and FNMR_All
60 mic_mic

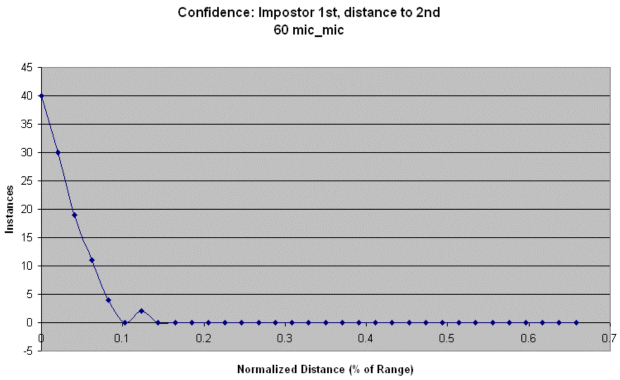
5)


FMR and FNMR_Zoom
60 mic_mic

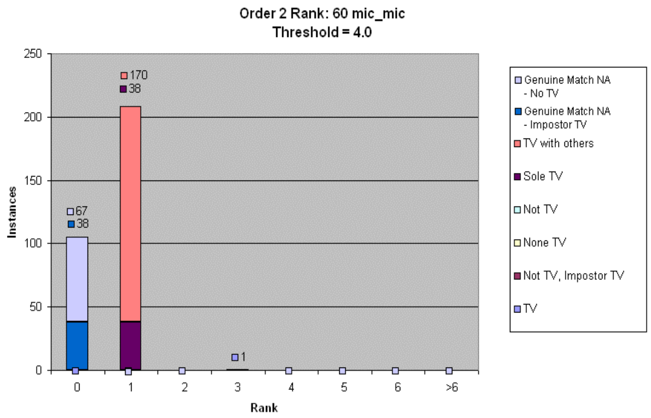
6)


Confidence: Genuine is 1st, distance to 2nd
60 mic_mic

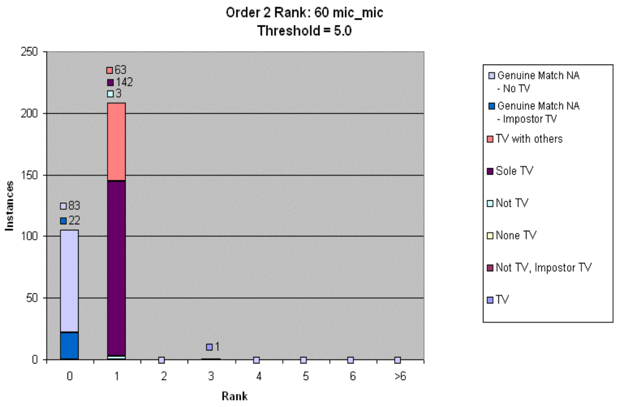
7)


Confidence: Impostor 1st, distance to 2nd
60 mic_mic

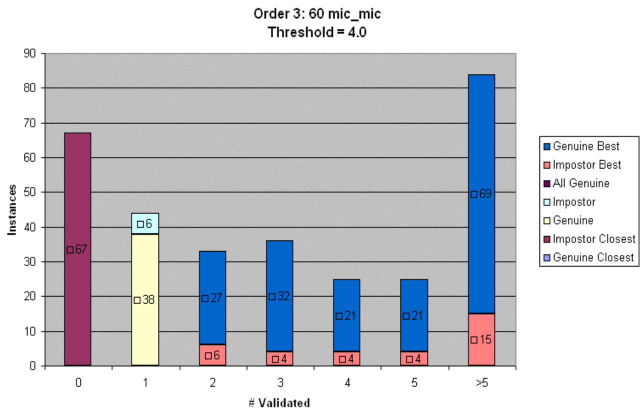
8)



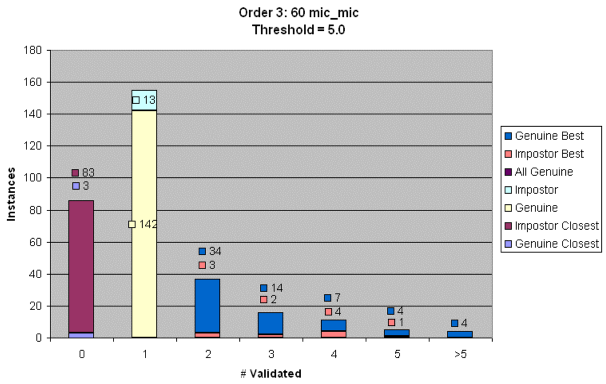
9)



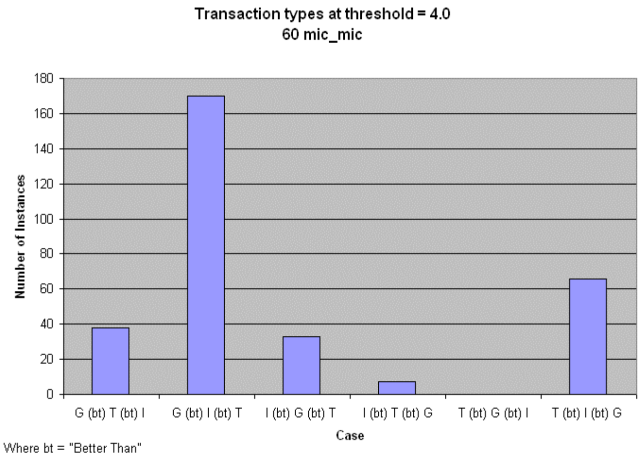
10)



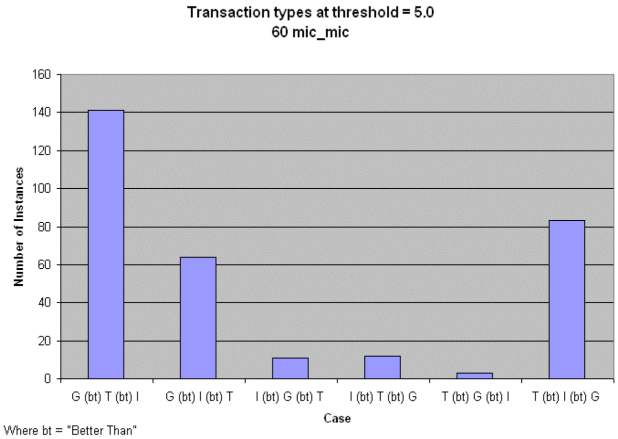
11)



12)



13) Where bt = "Better Than"



14) Where bt = "Better Than"

REFERENCES

[1] Joseph Atick. "Global And Mobile Identity Management: Business Processes And Technical Innovations To Ensure The Secure Flow Of Identities", Invited Presentation, Biometric Summit 2010, Miami, February 22-25, 2010.

[2] ISO SC 37 WD 29195, Technical Report on passenger processes for biometric recognition in automated border crossing systems.

[3] Dmitry O. Gorodnichy, "Multi-order Biometric Score Analysis Framework and its Application to Designing and Evaluating Biometric Systems for Access and Border Control". IEEE Workshop on Computational Intelligence in Biometrics and Identity Management, IEEE Symposium Series on Computational Intelligence – SSCI. April 11-15, 2011 - Paris, France

[4] D.O. Gorodnichy. "Evolution and evaluation of biometric systems". In IEEE Workshop on Applied Computational Intelligence in Biometrics, Proc. of IEEE Symposium: Computational Intelligence for Security and Defence Applications (CISDA), Ottawa, Canada, 2009.

[5] Dmitry O. Gorodnichy, "How To Conduct An All-Inclusive Performance Evaluation Of Your Biometric System", Invited Presentation, Biometric Summit, 2010 , Miami, February 22-25, 2010.

[6] D.O. Gorodnichy. "Multi-order analysis framework for comprehensive biometric performance evaluation". In Proceedings of SPIE Volume 7667: Conference on Defense, Security, and Sensing. - DS108: Biometric Technology for Human Identification track, Orlando, 5 - 9 April 2010.

[7] D.O. Gorodnichy, R. Hoshino. "Calibrated confidence scoring for biometric identification". In Proceedings of the NIST International Biometric Performance Conference (IBPC 2010), NIST Gaithersburg, March 2-4, 2010.

[8] D.O. Gorodnichy, R. Hoshino. "Score calibration for optimal biometric identification". Advances in Artificial Intelligence, 23rd Canadian Conference on Artificial Intelligence, Canadian, AI 2010, Ottawa, Canada, May 31 - June 2, 2010. Proceedings. Lecture Notes in Computer Science 6085 Springer 2010, ISBN 978-3-642-13058-8

[9] Alexey Khitrov, Francisco Ibarra. "Voice Biometrics For Law Enforcement: An Overview Of The World's First Large-Scale Automatic Voice Identification System", Invited Presentation, Biometric Summit, 2011, Miami, March 6-8, 2011

[10] http://www.itl.nist.gov/iad/mig/tests/spk/2006/index.html

[11] Alvin F. Martin. Speaker Databases and Evaluation. Stan Li (Editor), Encyclopedia of Biometrics, Elsevier Publisher, 2009.

[12] Martin, A. F., et al., "The DET Curve in Assessment of Detection Task Performance", *Proc. Eurospeech '97,* Rhodes, Greece, September 1997, Vol. 4, pp. 1899-1903

[13] ANSI INCITS 409.3-2005 Biometric Performance Testing and Reporting - Part 3: Scenario Testing and Reporting

[14] ISO/IEC SC 37 19795-2:2007 Biometric performance testing and reporting - Part 2: Testing methodologies for technology and scenario evaluation

[15] ISO/IEC SC 37 FCD 19795-5, Information Technology — Biometric Performance Testing and Reporting — Part 5: Grading scheme for Access Control Scenario Evaluation

[16] International Biometric Group. Biometric Performance Certification and test plan - http://www.biometricgroup.com/testing_and_evaluation.html

[17] J. L. Wayman, A. K. Jain, D. Maltoni, and D. Maio, editors. *Biometric Systems: Technology, Design and Performance Evaluation*. Springer, New York, 2005.

[18] http://www.agnitio.es

[19] CBET Portal : https://partners.drdc-rddc.gc.ca/css/Portfolios/Biometrics (Human ID Systems)/C-BET

[20] www.Nexus.gc.ca

[21] Dmitry O. Gorodnichy, Dave Bissessar, Elan Dubrofsky, Jonathon Lee. Analyzing the performance and risks of biometrics systems using Comprehensive Biometrics Evaluation Toolkit (C-BET), Justice Institute of British Columbia and the U.S. DHS's Center of Excellence VACCINE Workshop on "Visual Analytics for Public Safety Professionals", Sept. 20 -21, New Westminster, BC, 2010.