

Д. В. ДЫРМОВСКИЙ, С. Л. КОВАЛЬ

ОСОБЕННОСТИ ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА СОВРЕМЕННЫХ СИСТЕМ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Обоснованы требования к организации человеко-машинного интерфейса для современных систем автоматической и автоматизированной идентификации личности, основанных на анализе биометрических признаков.

Ключевые слова: человеко-машинный интерфейс, идентификация личности, биометрическая система идентификации, голосовая биометрия.

Введение. Системы автоматической идентификации личности (САИ) по биометрическим признакам получают все большее распространение. Они предназначены для решения задач учета и мониторинга неизвестных лиц, выполнения криминалистических идентификационных экспертиз. Рассмотрим особенности интерфейса и структуры САИ на примере использования динамических идентификационных признаков речевого сигнала.

Оптимальное представление результатов работы САИ. В современных САИ используется процедура обучения на больших базах биометрических данных совпадающих и различающихся личностей. Например, в обучающих базах речевых САИ содержатся файлы речи нескольких тысяч дикторов, записанных в разных условиях. Корпусной подход позволяет автоматически выбрать оптимальные правила и параметры идентификации, сопоставляющие скалярному расстоянию x между сравниваемыми речевыми файлами значения вероятности совпадения/различия дикторов. САИ сравнивает пары дикторов и по найденному для них x выдает вероятностный результат тождества/различия дикторов, практическая интерпре-

тация которого может различаться для разных задач. Например, для задачи верификации достаточно принимать решение „Да“—„Нет“.

Для решения задачи идентификации необходимы оценка точности предлагаемого решения и оценка его неопределенности. Характеристики речевого сигнала существенно зависят от свойств каналов звукозаписи и звукопередачи, состояния диктора, типа речевой коммуникации, сопутствующих помех и искажений и т.п. В силу этого создать представительные обучающие базы данных невозможно. Существующие учетные САИ [1—3] выдают результат только в виде ранжированного списка сравниваемых дикторов, что неприемлемо для единичных сравнений. Интерфейс САИ должен позволять пользователю выбрать оптимальную форму представления результатов и учесть характеристики конкретных сравниваемых дикторов, ориентируясь на случаи и множественного, и единичного сравнения. Ни одна из существующих прикладных и исследовательских САИ этим требованиям не удовлетворяет [1—6].

Известно много способов представления результатов работы САИ [7—9], однако выбор способа, оптимального для речевых систем, не очевиден. САИ проводит поиск целевого диктора в списке проверяемых. Предлагается представлять результат работы САИ в виде списка похожих дикторов (СП), полученного усечением списка всех проверяемых дикторов на основе предлагаемых показателей оценки работы системы. Будем исходить из того, что любая САИ для каждой пары целевой диктор—проверяемый диктор вычисляет наборы характеристических признаков и скалярное расстояние x между этими наборами. Используются показатели: $FRR(x)$ — False Rejection Rate — оценка вероятности ошибки 1-го рода: вероятность того, что файлы с речью целевого („своего“) диктора из списка проверяемых файлов не попадут в СП, если расстояние x от них до эталона будет больше соответствующего заданному значению FRR . $FAR(x)$ — False Acceptance Rate — оценка вероятности ошибки 2-го рода: вероятность того, что файлы с речью нецелевого („чужого“) диктора из списка проверяемых файлов попадут в СП, если расстояние x от них до эталона будет меньше соответствующего заданному значению FAR . $LR(x)$ — likelihood Ratio — оценка отношения правдоподобия для гипотез совпадения и различия дикторов при данном x между сравниваемыми файлами. $LR(x)$ рассчитывается как отношение вероятности отклонить „своего“ диктора при расстоянии между сравниваемыми дикторами больше данного x к вероятности принять „чужого“ диктора за своего при расстоянии между сравниваемыми файлами меньше x :

$$LR(x) = \frac{FRR(x)/(FRR(x) + FAR(x))}{FAR(x)/(FRR(x) + FAR(x))} = \frac{FRR(x)}{FAR(x)}. \quad (1)$$

Формула (1) имеет следующее толкование: вероятность верности нулевой гипотезы для интервала значений расстояния между дикторами больше данного x равна отношению доли попавших в этот интервал совпавших пар дикторов (т.е. $FRR(x)$) к общему числу пар дикторов, расстояние между которыми попало в этот интервал (т.е. $FRR(x)+FAR(x)$). Аналогично толкуется и знаменатель дроби в формуле.

Введем следующие понятия: P — общая вероятность совпадения сравниваемых дикторов; DET-график (Detection Error Trade-off) — график зависимости FRR от FAR .

Пример представления этих величин для системы автоматической идентификации VoiceNet приведен в таблице и на рис. 1 и 2. На рис. 1 представлены результаты идентификации САИ VoiceNet одного диктора (его данные заданы строкой в верхнем окне экрана) при сравнении с большим списком дикторов. Результат каждого сравнения указан в строках нижнего окна экрана. В каждой строке указан номер сравниваемого диктора в списке сравнения, идентификатор в базе данных, FRR , FAR , LR , имя секции базы данных, имя карточки диктора, имя звукового файла, тип источника звука. На рис. 2 приведен DET-график для результата сравнения двух дикторов в системе VoiceNet.

FRR, %	FAR, %	LR, у.е.	P
29,07	0,01	2907	0,999
7,20	0,10	72	0,986
1,07	0,50	2,1	0,68
1,00	0,56	1,78	0,64
0,67	1,00	0,67	0,40
0,53	5,00	0,11	0,1
0,50	6,74	0,07	0,07
0,10	33,44	0,003	0,003

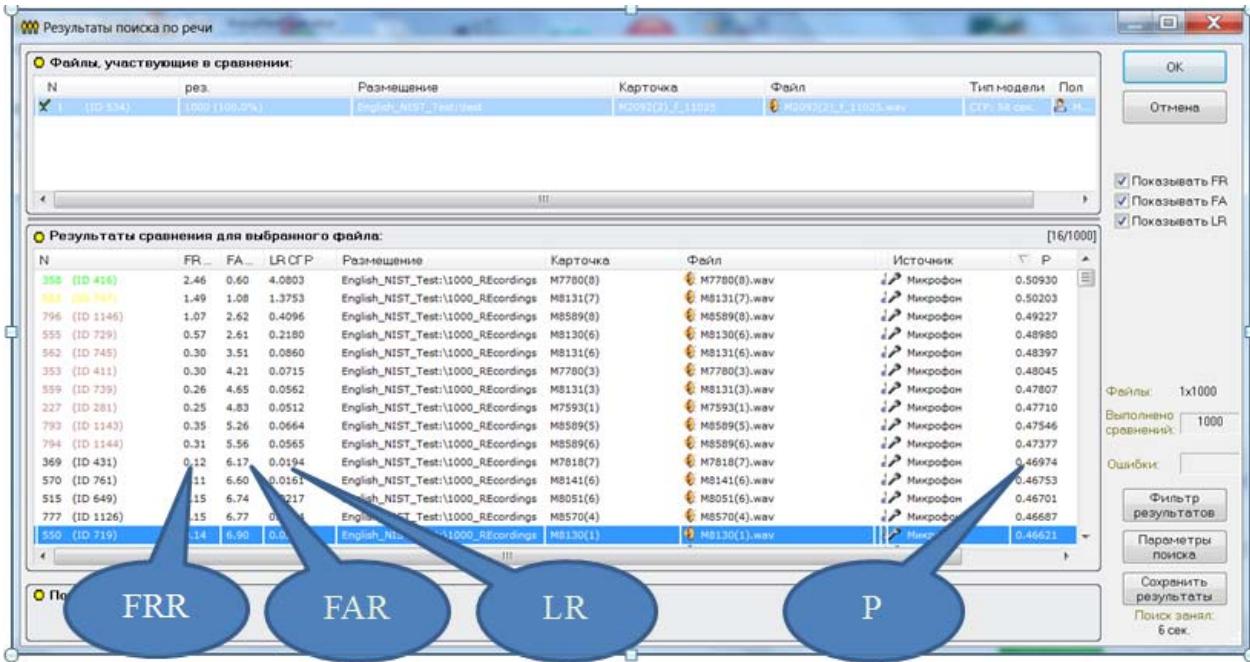


Рис 1

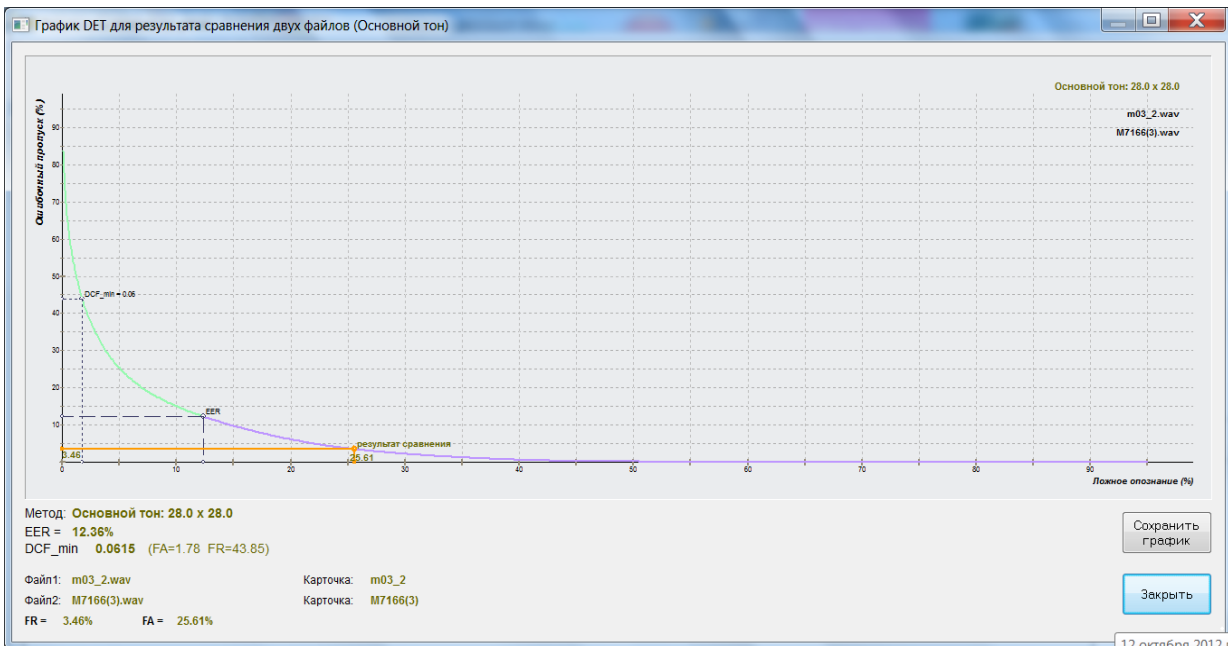


Рис. 2

Проиллюстрируем возможность использования этих показателей для конкретной САИ [10]. После сравнения диктора с данными базы из 100 000 голосов получим СП, ранжированный по степени сходства с данным диктором. Если ограничить рассматриваемый СП только дикторами, для которых $LR > 2907$, то диктор, находящийся в списке проверяемых, попадет в СП с вероятностью $\approx 70\%$, а с вероятностью $\approx 30\%$ — не попадет. Такой способ выбора порога обладает существенным преимуществом. Соответствующее выбранному порогу значение $FAR = 0,01\%$ означает, что в СП попадет не более 10 „чужих“ дикторов. Пользователь САИ за относительно малое время может проверить „ручными“, трудоемкими, экспертными средствами реальность тождества целевого диктора с этими 10 дикторами и выяснить, есть он действительно среди них или нет. Для 10 дикторов такая проверка на практике реализуема, а для большего числа дикторов уже трудноосуществима. Такой порог выгодно применять при проверке по большой базе неизвестных дикторов. Пропуск „своего“ диктора вероятен, но ввиду ограниченности ресурсов операторов системы по „ручной“ проверке СП выбор такого порога часто является единственной возможностью обнаружить в базе неизвестного искомого диктора.

При ограничении рассматриваемого СП только дикторами, для которых $LR > 0,67$, если целевой диктор есть в списке проверяемых, то он не попадет в СП с вероятностью всего $\approx 0,7\%$. Однако у такого выбора порога отсечки есть существенный недостаток. Соответствующее выбранному порогу значение $FAR = 1\%$ означает, что наиболее вероятно в СП попадет около 1000 „чужих“ дикторов, которые САИ сочтет близкими к целевому. Проверить „ручными“ средствами реальность тождества этих 1000 дикторов с целевым затруднительно. Тем не менее, такой порог выгодно применять при проверке по малой базе проверяемых дикторов. DET-график дает возможность выбора подходящих порогов отсечки для конкретной задачи с еще большей точностью, чем таблица.

Применение концепции доверительности данных к работе САИ. САИ применяются и в судебных экспертизах [4, 5, 11—14] при сравнении всего двух объектов. В этом случае результаты работы САИ целесообразно применять в рамках так называемого байесовского подхода [8, 15—17], объединяя данные исследований различных методов в единой формуле на основе значений LR по каждому из методов. Однако возможность применения статистических результатов обучения САИ к единичному случаю совершенно неочевидна. САИ вычисляет LR , что требует оценки неопределенности измерения [18]. Значение P можно обоснованно считать оценкой случайной величины, а в качестве параметров неопределенности результата оценки предлагается считать границы односторонних доверительных интервалов (ДИ) [19], которые определяются на основе подхода, близкого к методике NIST [20], которая использовалась для сравнительной оценки неопределенности результатов различных САИ. Нами рассматривается оценка неопределенности результатов работы отдельной САИ.

На этапе обучения САИ получает распределения частоты встречаемости расстояний x для пар совпадающих и различающихся дикторов. При решении задачи идентификации возможны два варианта: H_0 — сравниваемые дикторы совпадают и H_1 — различаются. Пусть $P(H_0|x)$ — апостериорная вероятность правильности гипотезы о совпадении дикторов. Тогда, согласно формуле Байеса:

$$P(H_0|x) = \frac{P(x|H_0)P(H_0)}{P(x|H_0)P(H_0) + P(x|H_1)P(H_1)}, \quad (2)$$

где $P(H_0)$ и $P(H_1)$ — априорные вероятности гипотез, $P(x|H_0)$ и $P(x|H_1)$ — вероятности получения x при верности каждой из гипотез. Значения $P(H_0)$ и $P(H_1)$ для простоты полагаются равными.

Апостериорная вероятность $P(H_0|x)$ моделируется сигмоидной функцией зависимости от x [21], оценка параметров которой проводится на обучающей базе данных. На рис. 3

приведен пример зависимости апостериорной вероятности $P(H_0|x)$ от x , полученной для конкретной САИ [10] на основе анализа гистограмм распределений x для совпадающих (эллипсы) и различающихся (крестики) дикторов в обучающей базе данных.

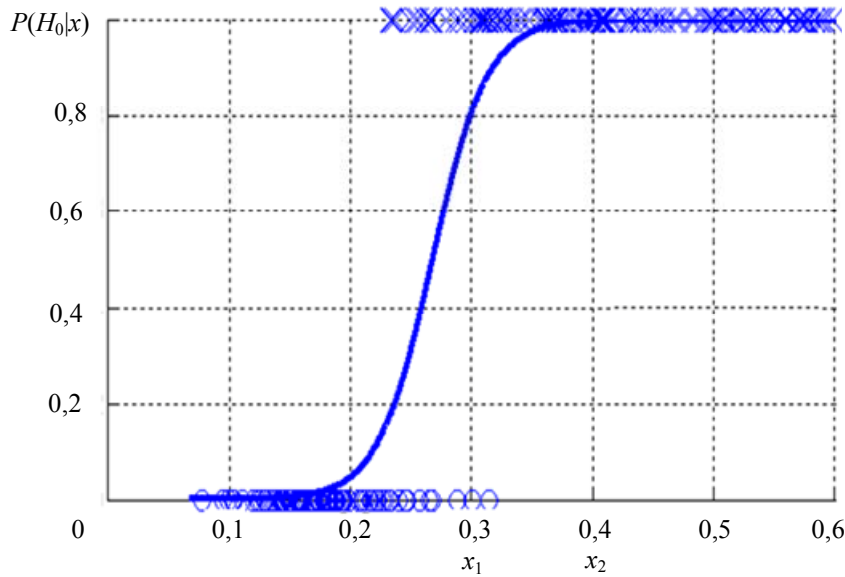


Рис. 3

Однако реальная форма распределения x для больших обучающих баз данных речевых САИ обычно далека от какого-либо стандартного типа. В силу этого при оценке ДИ для P предлагается использовать непараметрический bootstrap-метод, не требующий предположений о форме оцениваемого распределения [7, 20, 22]. Он позволяет оценивать распределение P для генеральной совокупности, используя только одну большую выборку. Пусть оценка апостериорной вероятности $\hat{P}_n = \hat{P}_n(H_0|x)$ получена для начальной выборки (X_1, \dots, X_n) . Чтобы оценить ДИ, в который с заданной вероятностью попадают значения $P(H_0|x)$, конструируем из начальной выборки большое количество других выборок, выбирая в произвольном порядке ее элементы „с возвратом“. Создадим наборы из B возможно повторяющихся элементов (X_1^*, \dots, X_n^*) и вычислим для них соответствующие значения $\hat{P}_n^*(b)$, $b = 1, \dots, B$. На основе оценок для bootstrap-выборок вычислим bootstrap-распределение \hat{P}_n^* : $G_*(p) = P\{\hat{P}_n^* \leq p\}$ — аналог распределения наборов обычных выборок из генеральной совокупности. Соответствующие процентилю этого распределения определяют квантили уровня значимости α и $1-\alpha$ $G_*^{-1}(\alpha) = \inf\{x : G_*(x) \geq \alpha\}$ и $G_*^{-1}(1-\alpha)$ как нижнюю и верхнюю границы $1-2\alpha$ ДИ для оценки $\hat{P}_n = \hat{P}_n(H_0|x)$ [22].

В качестве характеристик САИ выберем односторонний доверительный интервал (ОДИ). Верхний ОДИ:

$$P(-\infty < P(H_0|x) \leq \hat{P}_n^u) = \alpha, \quad (3)$$

где $\hat{P}_n^u = G_*^{-1}(1-\alpha)$, и нижний ОДИ:

$$P(\hat{P}_n^l \leq P(H_0|x) < \infty) = \alpha, \quad (4)$$

где $\hat{P}_n^l = G_*^{-1}(\alpha)$.

Для оценки результатов сравнений дикторов, близких к целевому, предлагается использовать нижнее значение ОДИ, а для дикторов, отличающихся от целевого — верхнее. Классический bootstrap-метод предполагает независимость элементов исходной выборки, что неверно для случая, когда сравниваются звуковые файлы одного и того же диктора. Для решения этой проблемы предлагается использовать subset bootstrap [23].

Экспериментальные результаты. На рис. 4 показана зависимость вероятности совпадения дикторов от расстояния между файлами (сплошная кривая) и кривая (пунктир) доверительных границ (ДГ), показывающая положение границ односторонних доверительных интервалов для $\alpha = 0,95$. Такой уровень доверительности означает, что отображаемая кривой доверительных границ $P(H_0|x)$ имеет значения „не хуже“ показанных на графике, по крайней мере, для 95% дикторов в обучающей базе данных. „Хуже“ и „лучше“ для пользователей САИ означает, что для заданного α положительное решение САИ о совпадении дикторов понимается только при условии $\hat{P}_n^l > 0,5$, а отрицательное только при выполнении условия $\hat{P}_n^u < 0,5$. Кривая ДГ для заданного уровня доверительности дает значение вероятности совпадения дикторов, минимальное — при условии принятия решения об их совпадении и максимальное — при условии принятии решения об их различии. При использовании ДГ возникает принципиально новая область возможных решений, в которой с заданным уровнем доверительности нельзя принять ни положительного, ни отрицательного решения ($LR=1$). На рис. 4 это зона для значений $x \in (0,26; 0,28)$.

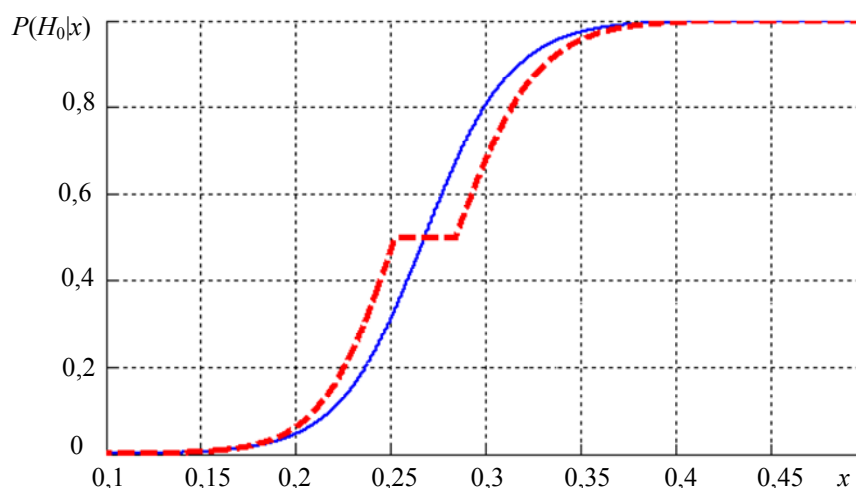


Рис. 4

Например, для данной САИ среднее значение апостериорной вероятности для значений $P(H_0|x_1) = 0,8$ и $P(H_0|x_2) = 0,98$ соответствует значениям кривой ДГ $\hat{P}_n^l(H_0 | x_1) = 0,67$ и $\hat{P}_n^l(H_0 | x_2) = 0,97$. На рис. 5 приведены результаты идентификации системы SISII (STC 2012), вероятностные результаты сравнения дикторов для двух речевых файлов, ДИ и ДГ (жирный шрифт).

На рис. 6 приведены ДГ для двух разных баз: микрофонные интервью в NIST SRE 2008 [24] и телефонные диалоги в аналоговых ТСОП в базе RuSTeN [25]. $P(H_0|x)$ САИ [19] для баз данных NIST SRE 2008 (1) и RuSTeN (2) и кривые доверительных границ для $P(H_0|x)$ для NIST SRE 2008 (3) и RuSTeN (4), вычисленные при $\alpha = 0,95$. Кривая ДГ показывает значения $P(H_0|x)$ для 95 % „лучших“, с точки зрения принимаемого идентификационного решения, дикторов в обучающей базе.

Идентификация

Файлы

Файл 1: C: 1341c_ch1_s_i.wav
16 бит; моно; 8000 Гц; 60.58 сек.;
речь не отсегментирована! чистая речь: 38.72 сек.;

Файл 2: A: 4e2c97_ch1_s_i_11025.wav
16 бит; моно; 11025 Гц; 46.67 сек.;
речь не отсегментирована! чистая речь: 28.08 сек.;

Уровень достоверности: 99%

Методы	FR [min,max], %	FA [min,max], %	LR [min,max]	P [min,max], %	P# [min,max], %	DET
<input checked="" type="checkbox"/> СФ	61.3 [57.7, 64.9]	2.4 [1.2, 3.5]	26.1 [17.6, 49.5]	96.3 [94.6, 98.0]	3.7 [2.0, 5.4]	DET
<input checked="" type="checkbox"/> ОТ	23.3 [20.2, 26.4]	10.5 [8.3, 12.8]	2.2 [1.7, 2.9]	68.9 [63.3, 74.2]	31.1 [25.8, 36.7]	DET
<input checked="" type="checkbox"/> СГР	8.9 [6.8, 10.9]	0.7 [0.1, 1.4]	11.8 [5.8, 75.2]	92.2 [85.2, 98.7]	7.8 [1.3, 14.8]	DET
<input checked="" type="checkbox"/> Общее ...	10.6 [8.4, 12.9]	0.4 [0.12, 0.9]	24.7 [10.6, 9 999.88]	96.1 [91.4, 99.99]	3.9 [0.01, 8.6]	DET

Заключение:

Идентификационные характеристики дикторов **совпадают** с уровнем достоверности более 99%.
Вероятность совпадения дикторов более 91.4%.

Рис. 5

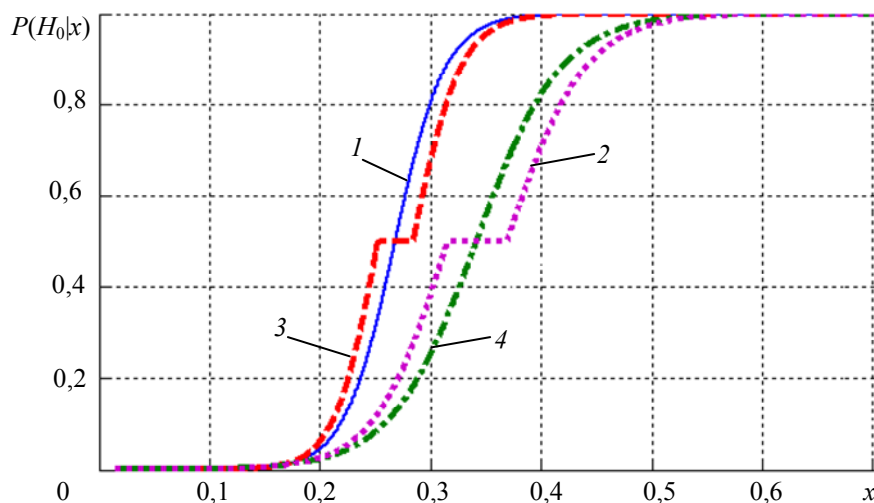


Рис. 6

Приведенные данные показывают высокую степень зависимости результатов работы САИ от типа обучающей базы данных, что требует наличия в пользовательском интерфейсе САИ настроек на свойства звуковых файлов.

Заключение. В работе предложены разработанные авторами, проверенные на практике подходы к организации пользовательского интерфейса САИ, ориентированные на решение задач мониторинга, учета и выполнения судебных экспертиз.

СПИСОК ЛИТЕРАТУРЫ

1. АПК „PhonoBase“ [Электронный ресурс]: <<http://www.sis-tss.ru/2010-06-23-20-32-10/2326-qphonobaseq.html>>.
2. [Электронный ресурс]: <http://www.agnitio.es/producto.php?id_producto=3>.
3. [Электронный ресурс]: <<http://www.loquendo.com/en/products/speaker-verification/>>.
4. Попов Н. Ф. и др. Идентификация лиц по фонограммам русской речи на автоматизированной системе „Диалект“. М., 1996.
5. Тимофеев И. Н. и др. Применение автоматизированной системы „Диалект“ на базе компьютерной речевой лаборатории CSL (США) при решении задач идентификации дикторов: Метод. рекомендации. ЭКЦ МВД РФ, 2000.
6. Martin A. F., Greenberg C. S. The NIST 2010 Speaker Recognition Evaluation // INTERSPEECH 2010. Makuhari, Chiba, Japan, 2010. P. 2726—2729.
7. Wu J. C., Martin A. F., Kacker R. N. Measures, Uncertainties, and Significance Test in Operational ROC Analysis // J. Res. NIST. 2011. Vol. 116, N 1. P. 517—537.
8. Campbell W. M. et al. Estimating and evaluating confidence for forensic speaker recognition // Proc. ICASSP2005. Philadelphia, PA, 2005.
9. Rose P. Technical forensic speaker recognition: Evaluation, types and testing of evidence // Computer Speech and Language. 2006. Vol. 20, N 2—3. P. 159—191.
10. Belykh I. N. et al. The speaker identification system for the NIST SRE 2010 // Informatics and its Applications. 2012. Vol. 6, N 1. P. 91—98.
11. Drygajlo A. Forensic automatic speaker recognition // IEEE Signal Processing Magazine. 2007. Vol. 24, N 2. P. 132—135.
12. Drygajlo A. Statistical Evaluation of Biometric Evidence in Forensic Automatic Speaker Recognition // IWCF 2009. Hague, Netherlands, 2009.
13. Interspeech 2008 special session “Forensic Speaker Recognition Traditional and Automatic Approaches“ [Электронный ресурс]: <<http://interspeech2008.forensic-voice-comparison.net>>.
14. Зубова П. И., Коваль С. Л. Методика экспертной идентификации дикторов по голосу и речи на основе комплексного анализа фонограмм // Теория и практика судебной экспертизы. 2007. Т. 3, № 7. С. 68—76.
15. Evett I., Buckleton J. Some aspects of the Bayesian approach for evidence evaluation // J. of Forensic Science Society. 1989. Vol. 29. P. 317—324.
16. Meuwly D., Drygajlo A. Forensic speaker recognition based on a Bayesian framework and Gaussian mixture modeling // Proc. „Odyssey“. 2001. P. 145—150.
17. Gonzalez-Rodriguez J. et al. Robust likelihood ratio estimation in Bayesian forensic speaker recognition // Proc. Eurospeech. 2003. P. 693—696.
18. Guide to the Expression of Uncertainty in Measurement. Geneva, ISO, 1993.
19. Koval S., Lokhanova A. Confidence Bounds Curves as a Tool for Evaluation of Automatic Speaker Recognition Results Uncertainty // Proc. 14th Intern. Conf. on Speech and Computer. SPECOM 2011. Kazan, 2011. P. 284—289.
20. Wu J., Martin A. F., Greenberg C. S., Kacker R. N. Measurement Uncertainties in Speaker Recognition Evaluation // NIST Publication. 2010. P. 7722.
21. Platt J. Probabilistic outputs for Support Vector Machines and comparisons to regularized likelihood methods // Advances in Large Margin Classifiers. Cambridge: MIT Press, 1999.
22. Bolle R.M. et al. Error Analysis of Pattern Recognition Systems: the Subsets Bootstrap // Computer Vision and Image Understanding. 2004. Vol. 93, N 1. P. 1—33.
23. Efron B., Tibshirani R. J. An Introduction to the Bootstrap. NY, 1993.
24. [Электронный ресурс]: <<http://www ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2011S11>>.
25. LDC 2006S34 [Электронный ресурс]: <www ldc.upenn.edu>.

- Дмитрий Викторович Дырмовский** — **Сведения об авторах**
филиал ООО „ЦРТ“, Москва; директор филиала; Санкт-Петербургский национальный исследовательский университет информационных технологий, кафедра речевых информационных систем; соискатель; E-mail: ddv@speechpro.com
- Сергей Львович Коваль** — канд. техн. наук, доцент; филиал ООО „ЦРТ“, Москва; главный эксперт; E-mail: koval@speechpro.com

Рекомендована кафедрой
речевых информационных систем

Поступила в редакцию
22.10.12 г.