

УДК 621.396, 004.934.8'1

**В.Л. Селетков,
М.В. Кузнецов**

ИДЕНТИФИКАЦИЯ ДИКТОРА ПО ПОРЯДКОВЫМ СТАТИСТИКАМ СПЕКТРОВ РЕЧЕВЫХ СИГНАЛОВ

Рассматривается контекстно-независимая идентификация диктора по порядковым статистикам совокупности текущих амплитудных спектров Фурье сегментов речевых цифровых сигналов спонтанной речи после удаления пауз. Метод спектральной идентификации в качестве информационного использует вектор значений порядковых статистик (выборочных квантилей) текущих амплитудных спектров речевых сигналов всей совокупности сегментов анализа (спектральный образ). Решение об идентичности дикторов (допуске или отказе) принимается по минимуму евклидовой метрики нормированного спектрального образа анализируемого речевого сигнала диктора и спектральных образов базы целевых абонентов. Приводятся предварительные экспериментальные оценки надежности идентификации дикторов по спектральным образам в целом без предварительной спектральной обработки и сепарации речевых сигналов.

Ключевые слова: верификация и идентификация диктора; база речевых сигналов целевых абонентов; кепстральное преобразование; порядковые статистики; выборочные квантили; выборочные децили; вероятность ошибочной идентификации; вероятность неверного допуска диктора; вероятность ошибочного отказа; вероятность пропуска диктора.

Розглядається контекстно-незалежна ідентифікація диктора за порядковими статистиками сукупності поточних амплітудних спектрів Фур'є сегментів мовних цифрових сигналів спонтанної мови після видалення пауз. Метод спектральної ідентифікації в якості інформаційного використовує вектор значень порядкових статистик (вибіркових квантилів) поточних амплітудних спектрів мовних сигналів всієї сукупності сегментів аналізу (спектральний образ). Рішення про ідентичність дикторів (допуск або відмову) приймається за мінімумом евклідової метрики нормованого спектрального образу аналізованого мовного сигналу диктора і спектральних образів бази цільових абонентів. Наводяться попередні експериментальні оцінки надійності ідентифікації дикторів за спектральними образами в цілому без попередньої спектральної обробки та сепарації мовних сигналів.

Ключові слова: верифікація та ідентифікація диктора; база мовних сигналів цільових абонентів; кепстральні перетворення; порядкові статистики; вибіркові квантили; вибіркові децили; ймовірність помилкової ідентифікації; ймовірність невірної допуску диктора; ймовірність помилкової відмови; ймовірність пропуску диктора.

The context-independent speaker identification by serial statistics of the current amplitude of Fourier spectra of segments of the digital speech signals of the spontaneous speech after the removal of pauses is considered. The method of spectral identification as an information one uses a vector of values of order statistics (sample quantiles) of the

current amplitude spectrum of the speech signal of the totality of segments analysis (spectral image). The decision about the identity of the speakers (admission or rejection) shall be based on the minimum Euclidean metric of the normalized spectral image of the analyzed speech signal of the speaker and spectral images of the database of target subscribers. Preliminary experimental evaluation of the reliability of the speaker identification by spectral images in whole without the prior spectral processing and separation of speech signals are stated.

Keywords: *verification and speaker identification; database of speech signals target subscribers; casterline conversion; order statistics; sample quantiles; sample deciles; the probability of mistaken identity; the probability of an incorrect access of the speaker; the probability of erroneous rejection; the probability of missing a speaker.*

Постановка задачі. Решение задач распознавания (верификации и идентификации) диктора по речевым сигналам находит широкое применение в криминалистике, радиоразведке и контрразведке, антитеррористическом мониторинге, при обеспечении безопасности и контроле доступа к информационным, вычислительным и другим ресурсам [1].

При использовании радио и телефонных каналов связи идентификация абонентов существенно усложняется искажениями и помехами в каналах связи. Поскольку в большинстве практически важных случаев фонетическое содержание сравниваемых речевых сигналов различно, значительный интерес представляют контекстно-независимые методы статистической идентификации речевых сигналов в условиях помех естественного и искусственного происхождения, когда условия регистрации входного речевого сигнала и речевых сигналов базы целевых дикторов существенно отличаются.

В большинстве систем распознавания для идентификации диктора используется метод кепстрального преобразования речевых сигналов на сегментах времени 10–20 мс. Получаемые наборы от 10 до 30 коэффициентов кепстра на сегменте отражают индивидуальные характеристики голосового источника и анатомии речевого тракта. В дополнение к коэффициентам кепстра используются также их первые и вторые разности по времени, поэтому на сегменте может регистрироваться от 30 до 90 значений. В общем случае это число должно быть умножено на количество сегментов в речевом высказывании, что приводит к размерностям в тысячи параметров.

Кроме того, низкая эффективность решения задачи статистической кепстральной идентификации речевых сигналов в условиях естественных помех каналов связи обусловлена, прежде всего, негауссовским характером статистических распределений мгновенных значений коэффициентов кепстра. В силу существенной нестабильности параметров последних используются сложные методы аппроксимации многомерной функции плотности распределения вероятностей в пространстве признаков взвешенной смесью нормальных распределений (GMM – Gauss Mixture Models) и методы опорных векторов (SVM – Support Vector Machines) не являются робастными.

В этих условиях значительный практический и теоретический интерес представляет разработка новых подходов и методов контекстно-независимой статистической идентификации дикторов (абонентов телефонных каналов связи) по индивидуальным особенностям порядковых статистик текущих амплитудных

спектров речевих сигналів на сегментах стаціонарності в умовах помах [2, 3].

Основная часть. Рассмотрим метод спектральной идентификации дикторов, в котором, после удаления пауз дискретных отсчетов спонтанной речи цифровым детектором активности, в качестве информационного используется вектор порядковых статистик (выборочных квантилей) текущих амплитудных спектров речевых сигналов всей совокупности сегментов анализа [2, 3]. Этот метод позволяет уменьшить влияние дестабилизирующих факторов на вероятность правильной идентификации в силу асимптотической нормальности, состоятельности, эффективности и робастности оценок квантилей по значениям порядковых статистик выборок случайных величин с непрерывными плотностями распределений мгновенных значений [4].

При реализации метода идентификации по порядковым статистикам спектра выполняют обычные взаимосвязанные операции аналого-цифровой обработки: низкочастотную фильтрацию спонтанной речи в полосе 4...20 кГц; стандартное 8...16-битовое аналого-цифровое преобразование с частотой дискретизации 8...40 кГц, удаление пауз в последовательности временных отсчетов цифровым детектором активности, разбиение полученной непрерывной последовательности отсчетов на сегменты спектрального анализа с одинаковым числом отсчетов, соответствующих длительности 20...30 мс речевого сигнала; формирование на каждом сегменте спектрального анализа цифровых отсчетов текущего амплитудного спектра на основе стандартного дискретного преобразования Фурье всех отсчетов цифрового речевого сигнала на сегменте стаціонарності [2, 3]. Таким образом, для 20–30 секундной длительности выборки речевого сигнала без пауз (соответствующего спонтанному речевому сигналу длительности порядка 40–120 с.) можно получить порядка 500–5000 реализаций отсчетов текущих амплитудных спектров сегментов исходного речевого сигнала. Для иллюстрации на рис. 1. приведен трехмерный график совокупности 4000 текущих амплитудных спектров сегментов анализа спонтанного речевого сигнала без пауз. Каждая отдельная реализация амплитудного спектра в диапазоне частот 50–3000 Гц интерпретируется как соответствующее сечение некоего непрерывного потенциального рельефа, отображающего изменение интенсивности спектральных компонент по времени (от сегмента к сегменту) и частоте спектра.

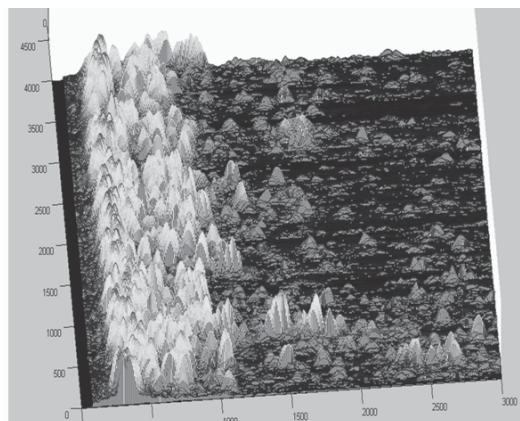


Рис. 1. Совокупность текущих амплитудных спектров сегментов речевого сигнала без пауз

Как и следовало ожидать, потенциальный рельеф всей совокупности реализаций амплитудных спектров речевых сигналов по сегментам анализа является многолепестковым (многопиковым) по частотам спектра. Параметры лепестка частоты основного тона, лепестков формантных и комбинационных частот, интенсивности и длительности пиков существенно изменяются в процессе разговора от группы в 10–50 сегментов к другой группе. В целом наблюдается сложный характер и большой динамический диапазон изменения значений амплитуд спектральных компонент речевого сигнала. Основная энергия по спектру речевого сигнала сосредоточена в полосе 50–1500 Гц, а энергетическими составляющими спектра на частотах выше 2000 Гц можно пренебречь [3].

При повторном произношении этого же контекста (например, для верификации диктора) в главных чертах можно ожидать воспроизведение непрерывного потенциального рельефа, отображающего изменение интенсивностей спектральных компонент по времени (от сегмента к сегменту) и частоте спектра, обусловленных фонетическими конструкциями и индивидуальными особенностями речи диктора. Для контекстно-независимых реализаций речи фиксируемый потенциальный рельеф будет только выборкой из возможных реализаций стохастического двумерного процесса.

При использовании радио- и телефонных каналов связи фиксируемый потенциальный рельеф будет искажаться за счет помех естественного и искусственного происхождения. Все это в совокупности обуславливают низкую помехозащищенность систем спектральной (кепстральной) идентификации дикторов в условиях шумов естественного и искусственного происхождения (шумового фона), помех в каналах передачи, регистрации и воспроизведения речевых сигналов, наличия аномальных помех (щелчки, всплески, реверберации и т.д.).

В этих сложных условиях повышение эффективности статистической идентификации диктора по индивидуальным особенностям спектральных компонент спонтанной речи в условиях негауссовости распределений отсчетов амплитудного спектра может быть получено на основе использования порядковых статистик отсчетов амплитудного спектра сегментов анализа для каждой спектральной частоты [2, 4].

После выполнения рассмотренных выше операций аналого-цифрового и спектрального преобразования принимаемого речевого сигнала для каждого дискретного значения частоты f_j текущего спектра Фурье может быть получена совокупность значений модулей $X_i(f_j), i = 1, n$ по всем n сегментам анализа. Эта совокупность статистически независимых случайных величин может считаться выборкой объема n (количество сегментов анализа) из генеральной совокупности с многомерной функцией распределения (ФР) $G[X(f_j)]$. В силу сложности статистических преобразований функция распределения $G[X(f_j)]$ не поддается разумному аналитическому описанию, пригодному для практического использования.

Вместе с тем, при достаточном значении объема выборки n , можно сравнительно просто (например, используя алгоритмы быстрой сортировки и формирования вариационного ряда $\tilde{X}_{1/n}(f_j), \dots, \tilde{X}_{n/n}(f_j)$), определить порядковые статистики (выборочные квантили) $X_p(f_j)$ для каждого дискретного значения частоты f_j текущего спектра Фурье [2, 4]. Для решения последующих задач статистической идентификации речевых сигналов наиболее пригодны

совокупности $X_{k/10}(f_j), k = \overline{1,9}$ выборочных децилей, выборочных пятилей $X_{k/20}(f_j), k = \overline{1,19}$ или выборочных процентилей $X_{k/100}(f_j), k = \overline{1,99}$, которые делят весь диапазон значений вариационного ряда соответственно на 10, 20 или 100 неравных по величине подинтервалов, в которые попадает одинаковое (не меньше десяти-двадцати) количество элементов ряда.

Анализируемые порядковые статистики $X_p(f_j)$ можно рассматривать как упорядоченную совокупность новых случайных величин (случайный вектор). Статистические характеристики порядковых статистик $X_p(f_j)$ вариационного ряда $\tilde{X}_{1/n}(f_j), \dots, \tilde{X}_{n/n}(f_j)$ существенно отличаются от статистических характеристик исходной выборки спектральных амплитуд речевого сигнала. Это обусловлено тем, что операция формирования элементов вариационного ряда по выборке независимых значений $X_i(f_j), i = \overline{1, n}$ непрерывной случайной величины $X(f_j)$ (ранжирование выборки) не является линейной операцией. Статистические характеристики случайных величин $X_p(f_j)$ конечно же зависят от вида функции распределения $G[X(f_j)]$ и объема выборки n . Вместе с тем, в пределе для $n \rightarrow \infty$, функция плотности распределения вероятностей (ПРВ) $g[X_p(f_j)]$ каждой порядковой статистики $X_p(f_j)$ асимптотически нормальна

$$g[X_p(f_j)] = c_p \exp \left[- \left| X_p(f_j) - \bar{X}_p(f_j) \right|^2 / 2\sigma_p^2 \right].$$

Математическое ожидание $\bar{X}_p(f_j)$ функции ПРВ $g[X_p(f_j)]$ удовлетворяет уравнению $G[\bar{X}_p(f_j)] = p$. Среднее квадратическое отклонение порядковой статистики $\sigma_p = \sqrt{p(1-p)/n / dG[\bar{X}_p(f_j)]}$ конечно, если статистика $X_p(f_j)$ генеральной совокупности единственна и дифференциал функции ПРВ $dG(x)$ существует и непрерывен в окрестности значения $x = X_p(f_j)$ [4].

При этом порядковые статистики $X_\alpha(f_j), X_\beta(f_j)$ являются коррелированными случайными величинами, хотя исходные случайные величины $X_i(f_j), i = \overline{1, n}$, из которых они формировались, были независимы. Коэффициенты корреляции $\rho_{\alpha,\beta}$ порядковых статистик $X_\alpha(f_j), X_\beta(f_j)$, описываемые выражениями $\rho_{\alpha,\beta} = \sqrt{\alpha(1-\beta) / \beta(1-\alpha)}$, постоянны и не зависят от типа исходного распределения $G[X(f_j)]$. В пределе, для $n \rightarrow \infty$, значения порядковых статистик $X_p(f_j)$ выборки стремятся к соответствующим значениям квантилей $\bar{X}_p(f_j)$ функций распределения $G[X(f_j)]$ генеральной совокупности [4].

Отметим, что при увеличении в M раз объема выборки $X_i(f_j), i = \overline{1, n}$ значения математических ожиданий $\bar{X}_\alpha(f_j), \bar{X}_\beta(f_j)$ порядковых статистик $X_\alpha(f_j), X_\beta(f_j)$ и их коэффициентов корреляции $\rho_{\alpha,\beta}$ не изменяются в силу особенностей соотношений связи их формирования. При этом значения дисперсий $\sigma_\alpha^2, \sigma_\beta^2$ статистик уменьшаются в M раз, что гарантирует асимптотическую эффективность и состоятельность несмещенных оценок квантилей [4].

Таким образом, при переходе от анализа выборки $X_i(f_j), i = \overline{1, n}$ случайных значений амплитудного спектра из генеральной совокупности, характеризуемой ФР $G[X(f_j)]$ (вид и параметры которых на практике обычно неизвестны), к анализу порядковых статистик (выборочных квантилей) $X_p(f_j)$ амплитудного спектра, существенно упрощается получение оценок параметров новых распределений для выборочных квантилей. Это обусловлено тем, что совокупность новых случайных величин – порядковых статистик $X_p(f_j)$, за исключением граничных значений соответствующего вариационного ряда $\tilde{X}_{1/n}(f_j), \tilde{X}_{n/n}(f_j)$, является асимптотически нормальной. При этом выборочные децили, пиатили и процентиля робастны к аномальным значениям спектральных компонент $X_i(f_j)$, которые расположены на “хвостах” соответствующего вариационного ряда $\tilde{X}_{1/n}(f_j), \tilde{X}_{n/n}(f_j)$.

Сформировав рассмотренным образом совокупность, например, выборочных децилей $X_{k/10}(f_j), k = \overline{1, 9}$, можно получить девять зависимостей $X_{k/10}(f), k = \overline{1, 9}$ выборочных децилей от спектральной частоты f . Для иллюстрации на рис. 2. приведены экспериментально полученные графики зависимостей 9, 8, 7, и 5-го выборочных децилей (90, 80, 70, и 50-го процентиля) в диапазоне спектральных частот 1500 Гц.

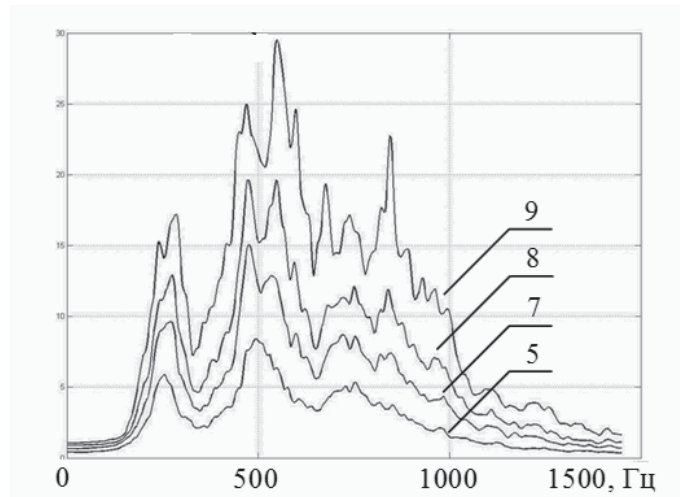


Рис. 2. Графики зависимостей 9,8,7, и 5-го выборочных децилей

В результате от анализа псевдослучайной совокупности значений $X_i(f_j)$ потенциального амплитудного рельефа всей совокупности реализаций амплитудных спектров речевых сигналов переходим к анализу статистически устойчивых зависимостей порядковых статистик $X_p(f)$ от частоты f спектра. При достаточной длительности анализируемого речевого сигнала, когда реализуются практически все основные фонетические конструкции речи, зависимости порядковых статистик $X_p(f)$ от частоты спектра отражают основные для данного сеанса индивидуальные особенности спектра речевого сигнала диктора независимо от контекста речи. Эти основные индивидуальные особенности спектра речевого сигнала диктора отражаются как в самом характере поведения зависимостей порядковых статистик $X_p(f)$ от частоты спектра – наличии локальных и глобальных экстремумов различной интенсивности для определенных значений частот, различной скорости изменения в поддиапазонах спектра и т.д., так и соотношением

значений зависимостей $X_\alpha(f_j), X_\beta(f_j)$ порядковых статистик. При этом индекс p порядковой статистики соответствующей зависимости $X_p(f)$ характеризует вероятность того, что значения случайных амплитуд спектра речевого сигнала во всем диапазоне не превосходят уровней $X_p(f)$.

Статистическая устойчивость и асимптотическая эффективность порядковых статистик $X_p(f)$ амплитудного спектра Фурье обеспечивают подобие зависимостей $X_p(f)$ при увеличении длительности анализируемого речевого сигнала (увеличении объема исходной выборки). В силу особенностей формирования зависимостей $X_p(f)$ они будут близкими для различных интервалов анализа речевого сигнала и от сеанса связи к сеансу. В тоже время для двух различных дикторов, независимо от контекстного содержания речи при достаточной длительности речевых сигналов, зависимости $X_p(f), Y_p(f)$ будут устойчиво отличаться при индивидуальных спектральных различиях речи. Для иллюстрации на рис. 3. приведены экспериментальные реализации 9, 8, 7 и 5-го децилей для шести различных дикторов в двух сеансах связи.

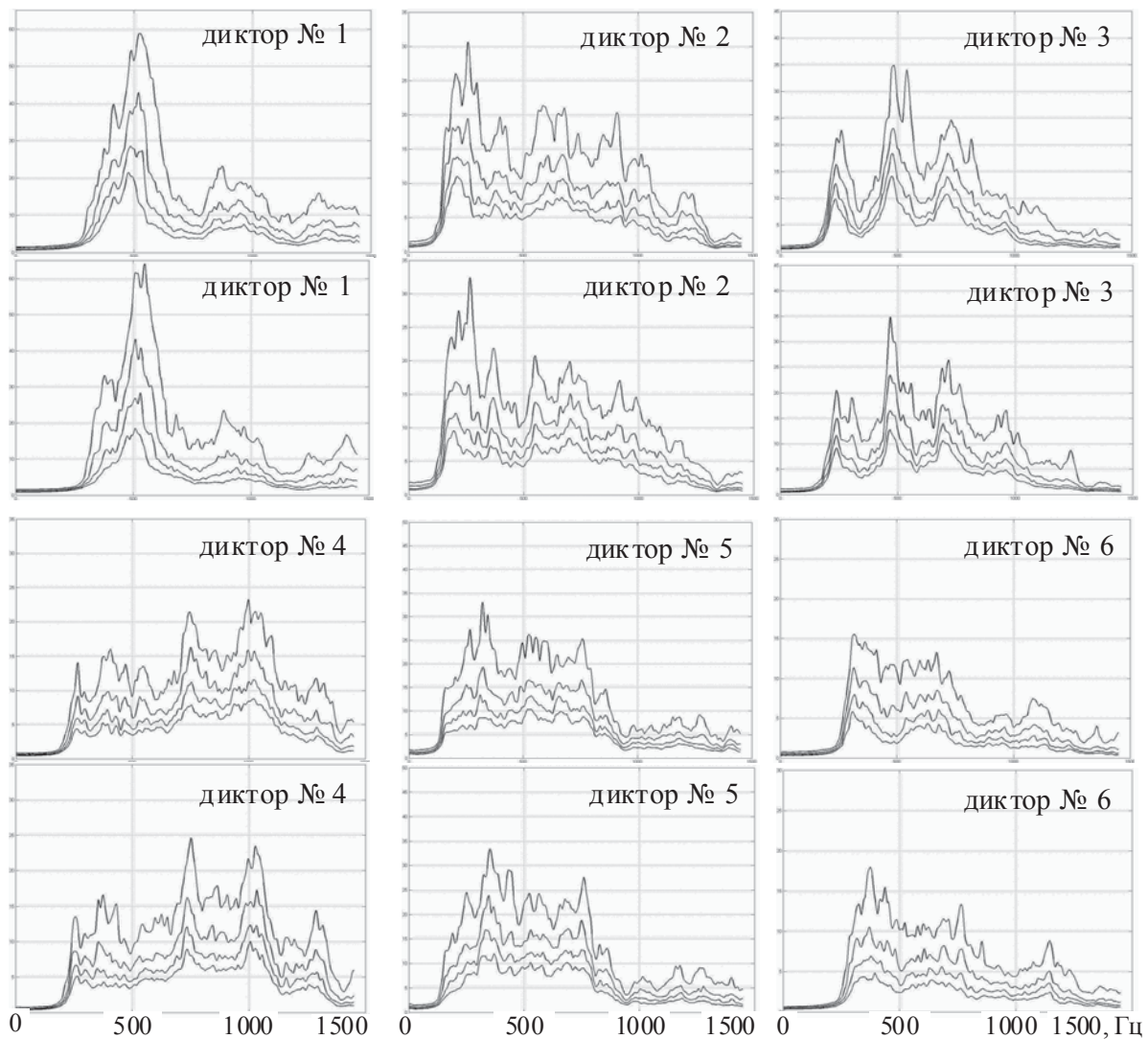


Рис. 3. Графики зависимостей 9, 8, 7, и 5 выборочных децилей для шести дикторов в двух сеансах связи

Уже при визуальном попарном сравнении этих экспериментальных зависимостей можно отметить подобие графиков для разных сеансов и существенные отличия зависимостей для различных дикторов.

Таким образом, для каждого k -го целевого абонента базы по достаточно длинным реализациям речевых сигналов (после удаления пауз детектором активности речи) можно сформировать составные векторы $\mathbf{X}_k = [X_p^{(k)}(f_j)]$ значений всех порядковых статистик $X_p^{(k)}(f)$ амплитудного спектра Фурье данного диктора. Этот составной вектор $\mathbf{X}_k = [X_p^{(k)}(f_j)]$ (называемый далее спектральным образом) является статистическим эквивалентом обработанного речевого сигнала, поскольку по значениям его компонент (порядковым статистикам спектра) методом Монте-Карло можно в частотной области сгенерировать многомерный псевдослучайный процесс, порядковые статистики которого будут статистически неразличимы с исходными статистиками.

В общем случае процедура идентификации диктора по его спектральному образу $\mathbf{Y} = [Y_p(f_j)]$ заключается в попарном его сравнении со спектральными образами $\mathbf{X}_k = [X_p^{(k)}(f_j)]$, $k = \overline{1, K}$ базы K целевых абонентов. Вектор разности $(\mathbf{X}_k - \mathbf{Y})$ спектральных образов абонентов является асимптотически нормальным с нулевым средним при совпадении абонентов и отличным от нуля средним для разных абонентов при одинаковых условиях регистрации соответствующих речевых сигналов. Решение об идентичности состоит в определении номера абонента по критерию минимума квадрата евклидовой нормы $\|\mathbf{X}_k - \gamma_k \mathbf{Y}\|^2 = (\mathbf{X}_k - \gamma_k \mathbf{Y})^* (\mathbf{X}_k - \gamma_k \mathbf{Y})$ разностей векторов спектральных образов базы целевых абонентов \mathbf{X}_k и входного нормированного образа диктора $\gamma_k \mathbf{Y}$:

$$\hat{k} = \arg \min_k \|\mathbf{X}_k - \gamma_k \mathbf{Y}\|^2 = \arg \min_k (\mathbf{X}_k - \gamma_k \mathbf{Y})^* (\mathbf{X}_k - \gamma_k \mathbf{Y}).$$

Значение $\gamma_k = \mathbf{X}_k^* \mathbf{Y} / \mathbf{Y}^* \mathbf{Y}$ коэффициента нормирования, учитывающего различия средней мощности идентифицируемых сигналов, соответствует минимуму значения квадрата $\|\mathbf{X}_k - \gamma \mathbf{Y}\|^2 = (\mathbf{X}_k - \gamma \mathbf{Y})^* (\mathbf{X}_k - \gamma \mathbf{Y})$ евклидовой нормы $\|\mathbf{X}_k - \gamma \mathbf{Y}\|$ разности образов $(\mathbf{X}_k - \gamma \mathbf{Y})$ для каждого k -го абонента:

$$\gamma_k = \arg \min_{\gamma} \|\mathbf{X}_k - \gamma \mathbf{Y}\|^2 = \arg \min_{\gamma} (\mathbf{X}_k - \gamma \mathbf{Y})^* (\mathbf{X}_k - \gamma \mathbf{Y}).$$

Основными показателями качества идентификации диктора являются вероятность ошибки первого рода (вероятность ошибочной идентификации или вероятность неверного допущения) и вероятность ошибки второго рода (вероятность ошибочного отказа или вероятность пропуска пользователя) при фиксированном пороговом значении U квадрата $\|\mathbf{X}_k - \gamma_k \mathbf{Y}\|^2$ евклидовой нормы. Экспериментальная оценка качества идентификации состоит в проведении статистического тестирования системы идентификации дикторов по совокупности $\sum_{k=1}^K M_k$ векторов $\mathbf{X}_k^{(m)}$, $k = \overline{1, K}$, $m = \overline{1, M_k}$ спектральных образов K абонентов в двух режимах.

В первом, для набора $\mathbf{X}_k^{(m)}, m = \overline{1, M_k}$ векторов спектральных образов k -го абонента, производится $M_k(M_k - 1)/2$ попарных сравнений (свой-свой) для определения такой же совокупности значений квадратов норм разностей $P_k^{(m,n)} = \|\mathbf{X}_k^{(m)} - \gamma_m \mathbf{X}_k^{(n)}\|^2, m \neq n$ векторов подобных спектральных образов k -го

абонента. Для всей совокупности $N_1 = \sum_{k=1}^K M_k(M_k - 1)/2$ величин $P_k^{(m,n)}, k = \overline{1, K}$

формируется вариационный ряд $P_{i/N_1}, i = \overline{1, N_1}$ их значений. Значения элементов $P_{i/N_1}, i = \overline{1, N_1}$ рассматриваемого ряда полагаются аргументами дискретной функции $\tilde{G}_1[P_k^{(m,n)}]$ из N_1 членов арифметической прогрессии $a_{n+1} = a_n + d = a_1 + d(n-1)$ с первым членом $a_1 = 1/N_1$ и разностью $d = 1/N_1$. Интерполяция (линейная, сплайнами) $G_1[P_k^{(m,n)}]$ рассматриваемой дискретной функции $\tilde{G}_1[P_k^{(m,n)}]$ на хвостах вариационного ряда $P_{i/N_1}, i = \overline{1, N_1}$ является достаточно точной аппроксимацией непрерывной функции распределения $\bar{G}_1[P_k^{(m,n)}]$ случайных значений $P_k^{(m,n)}$ квадратов норм невязок образов одного авторства генеральной совокупности. Интерполированная функция $G_1[P_k^{(m,n)}]$ позволяет для фиксированного значения порога U оценить вероятность правильной идентификации $\mathbf{P}_{1/1} = \mathbf{P}[P_k^{(m,n)} < U]$ или вероятность ошибки второго рода $\mathbf{P}_{0/1} = 1 - \mathbf{P}_{1/1}$ (вероятность ошибочного отказа или вероятность пропуска пользователя).

Во втором, для этого же набора $\mathbf{X}_k^{(m)}, m = \overline{1, M_k}, k = \overline{1, K}$ спектральных образов

производится $N_0 = \sum_{k=1}^{K-1} \sum_{m=k+1}^K M_m M_k$ попарных сравнений (чужой-чужой) для опре-

деления совокупности N_0 значений квадратов норм разностей $R_{k,r}^{(m,n)} = \|\mathbf{X}_k^{(m)} - \gamma_m \mathbf{X}_r^{(n)}\|^2, m = \overline{1, N_m}, n = \overline{1, N_n}, k = \overline{1, N_k}, r = \overline{k+1, N_r}$ векторов спектральных образов k -го и r -го абонентов. По всей совокупности N_0 величин $R_{k,r}^{(m,n)}$ формируется вариационный ряд $R_{i/N_0}, i = \overline{1, N_0}$ их значений. Значения элементов $R_{i/N_0}, i = \overline{1, N_0}$ рассматриваемого ряда полагаются аргументами дискретной функции $\tilde{G}_0[R_{k,r}^{(m,n)}]$ из N_0 членов арифметической прогрессии $b_{n+1} = b_n + c = b_1 + c(n-1)$ с первым членом $b_1 = 1/N_0$ и разностью $c = 1/N_0$. Интерполяция (линейная, сплайнами) $G_0[R_{k,r}^{(m,n)}]$ рассматриваемой дискретной функции $\tilde{G}_0[R_{k,r}^{(m,n)}]$ на хвостах вариационного ряда $R_{i/N_0}, i = \overline{1, N_0}$ является достаточно точной аппроксимацией непрерывной функции распределения $\bar{G}_0[R_{k,r}^{(m,n)}]$ случайных значений $P_k^{(m,n)}$ квадратов норм невязок различных по авторству образов генеральной совокупности. Интерполированная функция $G_0[R_{k,r}^{(m,n)}]$ позволяет для фиксированного значения порога U оценить вероятность ошибки первого рода $\mathbf{P}_{1/1} = \mathbf{P}[P_k^{(m,n)} < U]$ (вероятность ошибочной идентификации или вероятность неверного допущения) или вероятность правильного отказа $\mathbf{P}_{0/0} = 1 - \mathbf{P}_{1/0}$.

Для обеспечения надежности оценок вероятностей ошибочной идентификации $\mathbf{P}_{1/0}$ и ошибочного отказа $\mathbf{P}_{0/1}$ необходимо обеспечить статистическую значимость выборки тестируемых дикторов. В инженерной практике используется вытекающее из биномиального распределения эвристическое правило: для того, чтобы быть уверенным на 90 %, что полученная оценка верна, должно наблюдаться, по крайней мере, 30 ошибок. Для вероятности

пропуска пользователя в 1 % это означает, что должно быть не меньше 3000 испытаний, а 0,1 % – не меньше 30000 испытаний [1]. Например, если для $K=30$ абонентов сформировано по $M=15$ спектральных образов, то общее число попарных сравнений подобных спектральных образов составляет $N_1=3150$. При этом общее число попарных сравнений образов различных дикторов составляет $N_0=97875$. Приведенные оценки показывают, что при числе спектральных образов $M > 15$ для $K=30$ абонентов обеспечивается достаточно надежная оценка вероятности пропуска порядка 1 % и оценка вероятности ошибочной идентификации порядка 0,1 %.

Для технических приложений порог принятия решения U устанавливается для фиксированного значения вероятности ошибочной идентификации $P_{1/0}$ (как правило $P_{1/0}=10\%$). Для выбранного значения порога $U=U_{0,1}$ оценивается значение вероятности правильной идентификации $P_{1/1}(U_{0,1})$ или вероятность ошибочного отказа $P_{0/1}(U_{0,1})=1-P_{1/1}(U_{0,1})$. Однако при этом необходимо учитывать, что уменьшение порога U принятия решений для снижения вероятности ошибочной идентификации $P_{1/0}$ приводит к снижению значения вероятности правильной идентификации $P_{1/1}$ (росту вероятности ошибочного отказа $P_{0/1}=1-P_{1/1}$).

Системы идентификации диктора часто характеризуют значениями оценок вероятности правильной идентификации $P_{1/1}$ для фиксированных значений вероятности ошибочной $P_{1/0}$. Для рассматриваемого метода идентификации в табл. 1 приведены предварительные оценки значений вероятности правильной идентификации $P_{1/1}(P_{1/0})$ (в процентах) дикторов аудиокниг (АУДИО) и телефонных каналов связи (ТЛФ) для значений вероятности $P_{1/0}=[5\%,10\%,15\%,20\%,25\%,30\%]$.

Таблица 1

$P_{1/1}(P_{1/0})$	5%	10%	15%	20%	25%	30%
АУДИО	94%	95.5%	96.5%	97%	98%	98.5%
ТЛФ	70%	82.5%	86%	90.5%	94%	95%

Наиболее полное описание характеристик системы идентификации диктора дает оперативная (рабочая) характеристика – зависимость $P_{1/1}=F(P_{1/0})$ вероятности правильной идентификации $P_{1/1}$ от значений вероятности ошибочной $P_{1/0}$. Эквивалентной характеристикой является зависимость вероятности ошибочной идентификации $P_{1/0}$ от значений вероятности ошибочного отказа $P_{0/1}$ (DET–Detection Error Trade-off). DET-зависимость позволяет оценить показатель равной ошибки (EER – Equal Error Rate), определяющий вероятность ошибки распознавания $P_{EER}=P_{1/0}+P_{0/1}$ при условии равенства $P_{1/0}(P_{0/1})$ вероятности ошибочной идентификации (вероятности неверного допущения) и вероятности ошибочного отказа (вероятности пропуска пользователя). Показатель равной ошибки EER используют для сравнения различных методов распознавания диктора. По результатам тестирования на одной и той же базе данных, регулярно

проводимого в Национальном институте стандартов и технологий США (NIST), эти ошибки исследовательских систем находятся в диапазоне 3 %–5 %, так что суммарная ошибка распознавания равна удвоенной величине, т.е. 6 %–10 % [1].

Для рассматриваемого метода идентификации предварительные оценки значений вероятности ошибки распознавания составляют для дикторов аудиокниг (АУДИО) $P_{\text{EER}} = 5\%$ и телефонных каналов (ТЛФ) только $P_{\text{EER}} = 28\%$. Однако следует подчеркнуть, что рассматриваемая контекстно независимая идентификация дикторов по порядковым статистикам спектров спонтанных речевых сигналов (пятилям от 50 % до 95 % в диапазоне частот 50 Гц–1,5 кГц) проводилась без какой-либо предварительной спектрально-временной обработки и сепарации реализаций речевых сигналов по языку и полу абонента, качеству восприятия речи на слух и уровню мощности сигнала.

Дальнейшее повышение эффективности рассмотренного метода статистической идентификации можно обеспечить за счет многоканальной частотной обработки с учетом индивидуального характера изменения зависимостей выборочных квантилей от частоты, идентификации зависимостей по каждой порядковой статистике с мажоритарным правилом принятия решения об идентичности сигналов в целом и использования двух пороговых правил принятия решения об идентичности зависимостей порядковых статистик для снижения вероятности ошибочной идентификации.

В целом рассмотренный метод статистической идентификации по порядковым статистикам амплитудных спектров сигналов [2, 3] представляет определенный интерес для решения практических задач распознавания дикторов, а также идентификации источников акустических, гидроакустических и шумоподобных широкополосных сигналов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Распознавание личности по голосу : аналитический обзор / В.Н. Сорокин, В.В. Вьюгин, А.А. Тананикин // Электронный научный журнал ИППИ РАН – 2012. – № 1, том 12, С. 1–30 [Электронный ресурс] Режим доступа до журн. : [http : // www.jip.ru/2012/1-30-2012.pdf](http://www.jip.ru/2012/1-30-2012.pdf).
2. Вариант спектральной идентификации речевых сигналов / В.Л.Селетков // Изв. вузов. Радиоэлектроника. – 2006. – 49, N 3–4 [ч. 2]. – С. 56–61.
3. Пат. 80587 Украина, МПК 2006 G10L 15/00, G10L 17/00. Способ спектральной идентификации речевого сигнала / М.В. Кузнецов, В.Л. Селетков ; заявители и владельцы патента М.В. Кузнецов, В.Л. Селетков. – № а 200505554 ; заявл. 10.06.2005; опубл. 15.12.2006, Бюлл. № 12.
4. Дейвид Г. Порядковые статистики / Г. Дейвид ; пер. с англ. В. Егорова, В. Невзорова. – М. : Наука, 1979. – 335 с.

Отримано 30.09.2014

Рецензент Рибальський О.В., доктор технічних наук, професор.