

Метод идентификации речевых сигналов на основе порядковых статистик амплитудных спектров

Кузнецов М.В., Киев, Украина



• речевой сигнал • идентификация речевого сигнала • порядковые статистики спектра • голосовой статистический образ.

В статье представлен новый метод идентификации речевых сигналов, основанный на применении зависимостей от частоты спектра сигнала квантилей вариационных рядов — порядковых статистик спектра Фурье речевого сигнала. В силу нормализации распределения значений спектра в составе выборочных квантилей, достигается повышение эффективности идентификации речевых сигналов известными методами статистической проверки гипотез по критерию Неймана-Пирсона. Метод идентификации приобретает свойства устойчивости к изменениям как статистических характеристик речи, так и статистических характеристик каналов передачи и воспроизведения речевых данных. Представлены данные, характеризующие эффективность разработанного метода.

speech signal • speech signal identification • spectrum's order statistics • statistical pattern of the speech signal

This article is about the newest speaker identification method using by speech. This method is based on application of variational series fractals depending on speech signal spectrum frequency — the order statistics of speech signal spectrum. Method is robust to abnormal noise, natural interference, man-made noise effects in channels of transmission, recording and playback of speech signals. The method efficiency data is submitted.

Постановка проблемы

В настоящее время в полной мере нерешённой и актуальной остаётся задача идентификации диктора по его речевому сигналу. Идентификация (англ. identification — отождествление) в общем случае — это процедура установления тождественности процессов, объектов, персон. Процедура идентификации состоит из двух этапов: сопоставления совокупностей отдельных общих и индивидуальных признаков идентифицируемых объектов и принятия решения об идентичности по критериям проверки конкурирующих гипотез.

Трудность решения задачи обусловлена природой речевого сигнала, который является случайным нестационарным процессом, с медленно меняющейся дисперсией и сложной формой текущего спектра сигнала. Вследствие этого многомерные плотности распределения вероятностей параметров текущих спектров Фурье речевого сигнала являются негауссовскими и многомодальными. Это значительно усложняет применение извест-



ных статистических критериев согласия, ориентированных на статистическую идентификацию нормально распределённых случайных величин и существенно снижает вероятность правильной идентификации. Вышеперечисленные обстоятельства обуславливают актуальность разработки новых и совершенствования существующих методов спектральной идентификации¹ речевых сигналов.

Цель: ознакомление специалистов в сфере распознавания речевых сигналов и признакового описания речевых сигналов с новым разработанным методом идентификации дикторов по речи, основанным на сопоставлении порядковых статистик спектра Фурье речевого сигнала.

1. Спектральная идентификация речевых сигналов

Основой анализа речевых сигналов и базисом многочисленных методов исследования речи является математический аппарат преобразования Фурье. Спектральный анализ позволяет охарактеризовать частотный состав измеряемого процесса. Теоретически допускается, что исследуемый сигнал квазистационарный на конечном интервале и рассматривается на бесконечно протяжённом участке времени или на конечном интервале, который в большинстве случаев не может быть очень большим, чтобы обеспечить приемлемую скорость обработки. Такой подход подразумевает использование в спектральном анализе статистических методов и даёт возможность говорить о полученном результате не как о точном представлении частотного состава исследуемого процесса, а лишь как о некоторой его оценке [1].

В современных системах идентификации с применением спектральных методов используются цифровые способы обработки речевых сигналов. Если представить последовательность отсчётов речевого сигнала $x(t)$ на сегменте анализа в виде: $x[n]; n=0,1,\dots,N-1$, то спектральную плотность Фурье этой реализации можно записать в виде преобразования Фурье последовательности $x[n]; n=0,1,\dots,N-1$:

$$X_F[n] = \sum_{k=0}^{k=N-1} x[k] \exp(-j \frac{2\pi}{N} kn) = XC[n] - jXS[n],$$

где $XC[n]; n=0,1,\dots,N-1$ — отсчёты чётной реальной части спектральной плотности Фурье; $XS[n]; n=0,1,\dots,N-1$ — отсчёты нечётной мнимой части спектральной плотности Фурье [1].

Для идентификации анализируемого речевого сигнала по всей совокупности отсчётов спектральных плотностей Фурье определяют статистические характеристики параметров спектра Фурье для каждой частоты спектра речевого сигнала по всем сегментам анализа. Полученные статистические характеристики параметров спектра сравнивают по известным критериям согласия с соответствующими статистическими характеристиками параметров спектров образцовых речевых реализаций базы данных [2].

В ситуации, когда изначально характеристики сигнала неизвестны, трудно сказать, какая из спектральных характеристик с большей степенью достоверности отображает истинный спектр анализируемого сигнала. Такие неопределённости, часто возникающие на практике, показывают субъективный характер спектрального анализа.

¹ Спектральная идентификация — идентификация речевых сигналов с использованием спектральных методов анализа речевых сигналов, методов, основанных на исследовании параметров сигнала по его спектру.



Кузнецов М.В.

Метод идентификации речевых сигналов на основе порядковых статистик амплитудных спектров

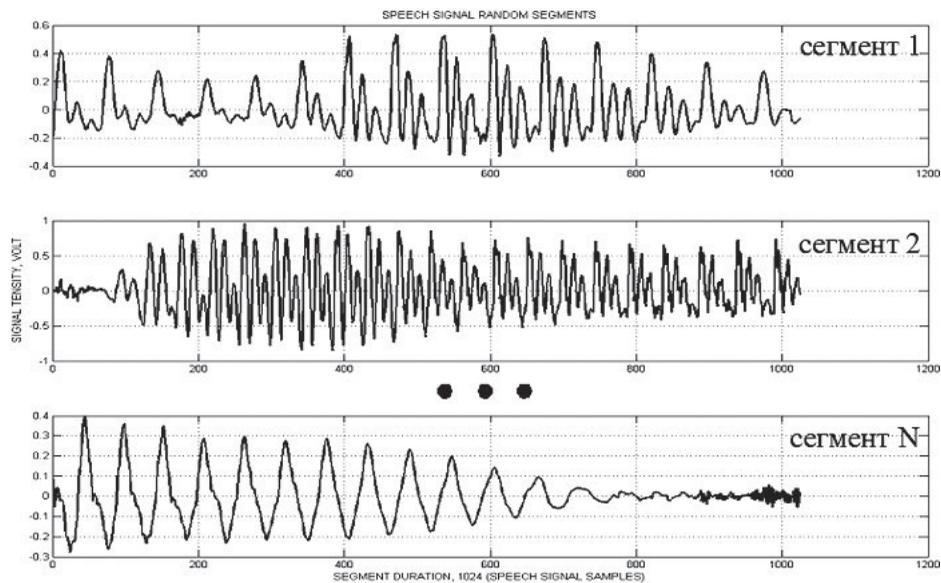


Рис. 1. Выборочные сегменты речевого сигнала, диктор № 1

В силу нестационарного характера речевого сигнала (рис. 1), реализации выборочных амплитудных спектров по сегментам анализа для одного и того же диктора существенно различаются (рис. 2).

На рис. 1 представлены выборочные сегменты речевого сигнала диктора № 1. По оси абсцисс отложена длительность сегмента в отсчётах, которая составляет 1024 отсчёта сигнала, что для частоты дискретизации 8820 Гц составляет 116 мс. По оси ординат — амплитуда напряжения сигнала в вольтах.

На рис. 2 представлены амплитудные спектры Фурье выборочных сегментов (рис. 1). По оси абсцисс отложены значения частоты спектра в Гц, по оси ординат — амплитуда напряжения сигнала в вольтах.

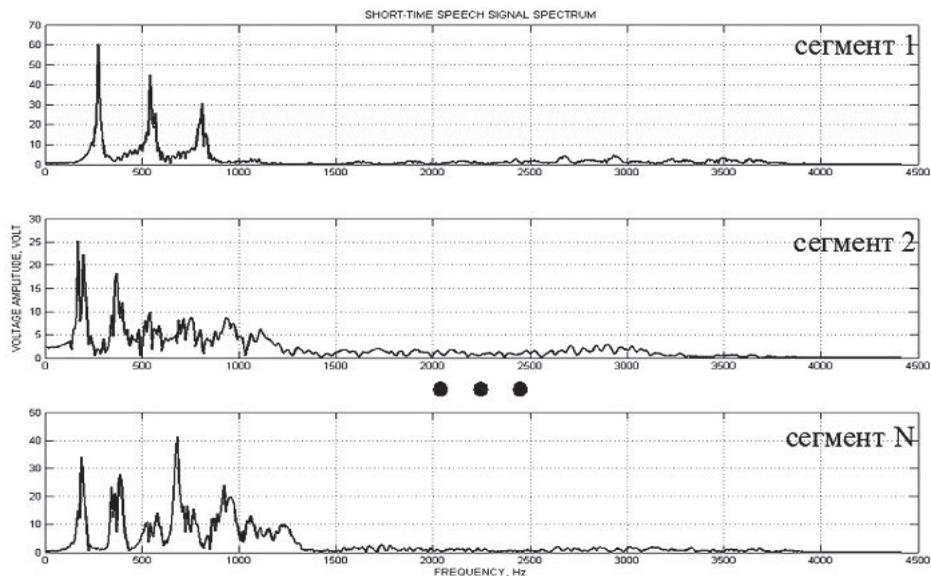


Рис. 2. Амплитудные спектры Фурье на выборочных сегментах анализа речевого сигнала, диктор № 1



На практике имеют место существенные отклонения распределений значений сигнала и амплитуд спектра выборочных реализаций речевого сигнала от нормальных распределений. Наблюдается нестационарный характер и большой динамический диапазон изменения значений интенсивностей спектральных компонент речевого сигнала.

Для решения задач спектральной идентификации речевых сигналов применение усреднённых на больших периодах времени спектров речи [3] приводит к нивелированию индивидуализирующих параметров речевого сигнала, которые отображаются на спектре сигнала. И, как следствие, приводит к снижению вероятности правильной идентификации [4, 5].

В этом случае **повышение эффективности** систем и методов спектральной идентификации речевых сигналов может быть достигнуто за счёт метода, основанного на применении зависимостей от частоты спектра сигнала квантилей вариационных рядов — порядковых статистик спектра Фурье речевого сигнала.

2. Порядковые статистики спектра речевого сигнала

Порядковые статистики амплитудного спектра, в силу особенностей своего формирования [6], робастны к рассмотренным выше аномальным дестабилизирующим факторам.

Вариационный ряд $[\hat{x}_{i/n}]$ значений $[x_i]$ частоты амплитудного спектра Фурье: $\hat{x}_{1/n} \leq \hat{x}_{2/n} \leq \dots \leq \hat{x}_{i/n} \leq \dots \leq \hat{x}_{n-1/n} \leq \hat{x}_{n/n}$ содержит всю информацию об исходной выборке $[x_i]$. Поэтому он называется **тривиальной достаточной статистикой** всей совокупности независимых случайных величин $[x_i]$. Значения $x_i, i=1 \dots n$ реализации $[x_i]$ случайны и в общем случае непредсказуемы с разумной для практики точностью. Но если от выборки $X_k = [x_i]_k$ к выборке $X_m = [x_i]_m$, от одной реализации вектора X_k к другой реализации X_m , значения $x_i, i=1 \dots n$ являются случайными величинами и независимы между собой, то в значениях элементов вариационного ряда $[\hat{x}_{i/n}]$ начинают проявляться статистические связи и элементы устойчивости предсказания значений этих элементов с приемлемой для практики точностью. Таким образом, вектор квантилей вариационного ряда $[\hat{x}_{i/n}]$ является асимптотически (при $n \geq 30$) гауссовским вектором при произвольном законе распределения $G(x)$ элементов генеральной совокупности, если производная $g(x)$ плотности распределения вероятностей (ПРВ) непрерывна [6].

В силу случайного характера речевых процессов, значения квантилей спектров Фурье \hat{x}_α , получивших в литературе общее название **порядковых статистик**, являются случайными и меняются независимо от выборки к выборке. В общем случае порядковые статистики \hat{x}_α можно рассматривать как упорядоченную совокупность новых случайных величин (случайный вектор), статистические характеристики которой зависят от вида функции $g(x)$ и объёма выборки n . Вместе с тем в пределе (при $n \geq 30$) распределение каждой порядковой статистики \hat{x}_α вариационного ряда $[\hat{x}_{i/n}]$ асимптотически нормально.

Таким образом, при переходе от анализа выборки $[x_i]$ из генеральной совокупности, характеризуемой $g(x)$ (ПРВ) и функцией распределения $G(x)$ (вид



и параметры которых на практике обычно неизвестны), к анализу вариационного ряда $[\hat{x}_{i/n}]$ существенно **упрощается** получение оценок параметров новых распределений. Это обусловлено тем, что за исключением экстремальных значений $\hat{x}_{1/n}, \hat{x}_{n/n}$ совокупность новых случайных величин является асимптотически нормальной [6].

3. Статистический вектор квантилей. Робастность метода

В представленном методе спектральной идентификации использовались речевые сигналы, которые представляют собой русскоязычные и англоязычные тексты в виде аудиокниг (производство «Сидиком», Украина). Характер речевого материала неоднородный. При чтении дикторы меняют интонации, пытаются имитировать голоса детей и пожилых людей с целью передачи настроения и смысла произведений. В исследовании применялись записи 15 дикторов. Исходный речевой материал подвергался предварительной обработке: понижению частоты дискретизации с 44100 Гц до 8820 Гц (редактор аудио-файлов CoolEdit Pro v. 2.0), удалению пауз в речи, сегментации (размер сегмента — 1024 отсчёта сигнала, что для частоты дискретизации 8820 Гц составляет 116 мс). Посегментно проводилось вычисление быстрого преобразования Фурье (БПФ), формирование зависимостей квантилей, с последующим построением статистического вектора квантилей речевого сигнала (голосового статистического образа) и проведением идентификационного сопоставления (рис. 3) значений статистического вектора квантилей исследуемых речевых сигналов со значениями статистических векторов квантилей речевых сигналов базы данных, т.е. сигналов, принадлежность которых источнику установлена достоверно. Статистический вектор квантилей насчитывает 100 сегментов анализа.

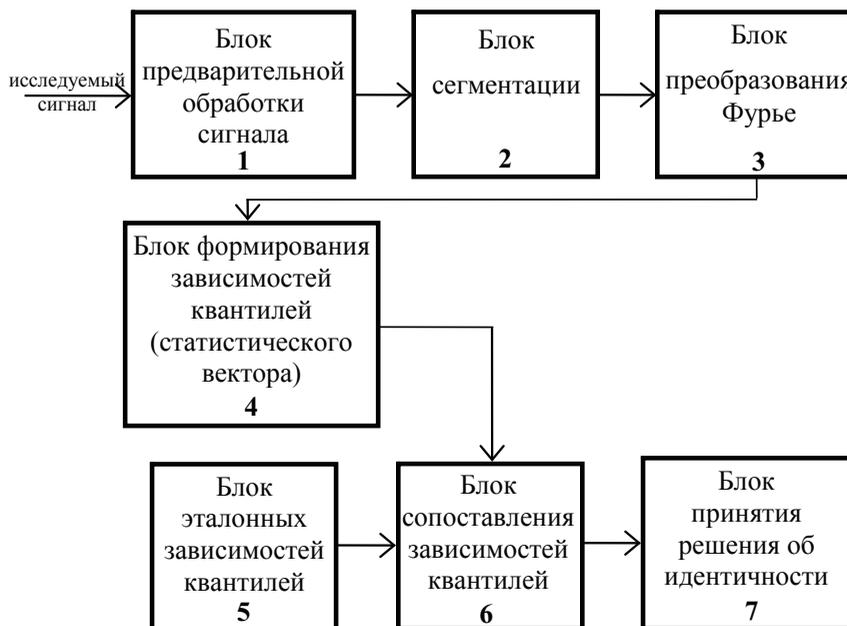


Рис. 3. Структурная схема метода спектральной идентификации речевых сигналов на основе порядковых статистик спектра

При формировании голосового статистического образа на основе порядковых статистик спектра речевого сигнала **не происходит усреднения** значений спектра речевого сигнала. Следовательно, сохраняются характерные для диктора особенности речевого сигнала, которые отображаются на спектре сигнала и на квантильных зависимостях.

Поскольку индивидуальность строения речеобразующего тракта человека — очевидный факт, речевой сигнал в разработанном методе рассматривается как целостный процесс, содержащий в себе все характеризующие параметры. По алгоритму разработанного метода не требуется дополнительно исследовать составляющие психофизического состояния диктора, смыслового наполнения речи, проводить акустический и лингвистический анализ сигналов.

С целью продемонстрировать достигаемую степень различий между голосовыми статистическими образами разных дикторов, на рис. 4 приведены голосовые статистические образы четырёх дикторов: «4а» — диктор № 1 женщина, язык английский; «4б» — диктор № 3 мужчина, язык русский; «4в» — диктор № 5 мужчина, язык английский; «4г» — диктор № 7 женщина, язык русский.

По оси абсцисс отложены значения частоты спектра сигнала, выведены значения до 1500 Гц, по оси ординат — амплитуда напряжения в вольтах, значения которой отображают величину значений квантилей. Наглядно оценить степень различия между полученными голосовыми статистическими образами разных дикторов можно по местоположению на оси частот максимальных значений зависимостей квантилей и по ширине полосы занимаемых ими частот, а также по величине энергетической составляющей зависимостей квантилей спектра (рис. 4).

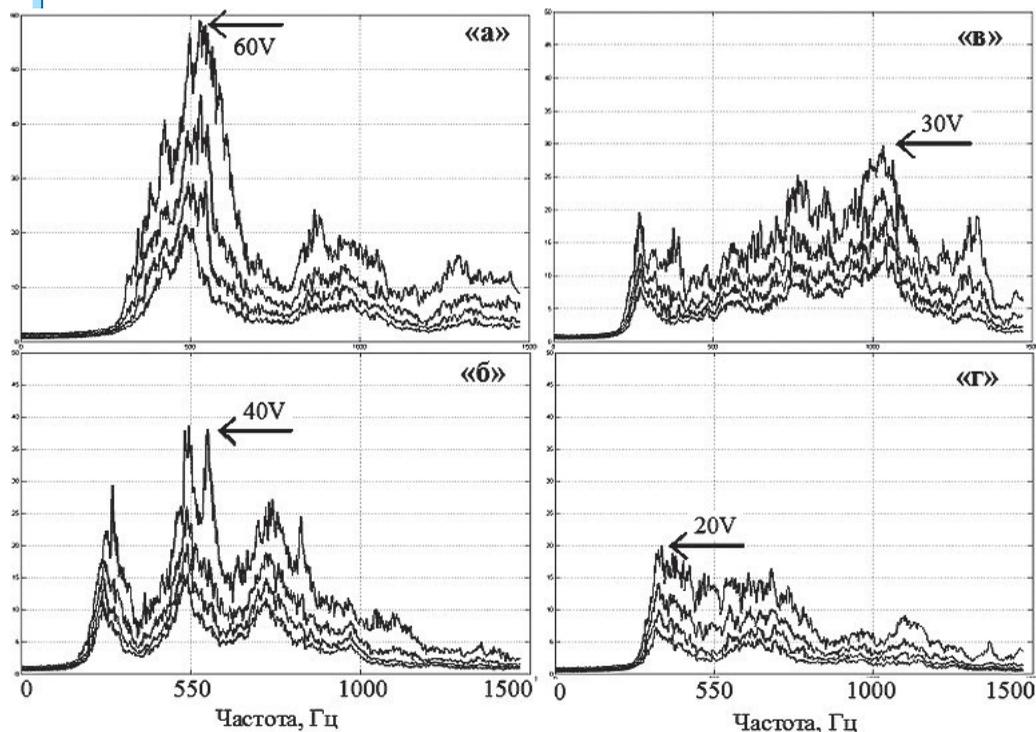


Рис. 4. Выборочные голосовые статистические образы исследуемых речевых сигналов, дикторы № 1 (4а); № 3 (4б); № 5 (4в); № 7 (4г)

Для решения задачи идентификации можно менять уровень **детализации образа**, с целью изменения требуемого объема информационного наполнения образа. Реализован уход от сопоставления параметров многомерных суммарных **ПРВ** к сопоставлению **числовых** значений векторов квантилей

Кузнецов М.В.

Метод идентификации речевых сигналов на основе порядковых статистик амплитудных спектров

вариационного ряда. Особенности алгоритма формирования порядковых статистик [6] спектра Фурье речевого сигнала позволяет **смещать** anomalно слабые и anomalно мощные значения зависимостей в области, которые не будут задействованы в идентификации. Это свойство позволяет нивелировать вредное влияние импульсных помех и шумов различного происхождения при анализе речевых сигналов без нарушения статистического наполнения сигнала и потерь информационной составляющей сигнала (рис. 5).

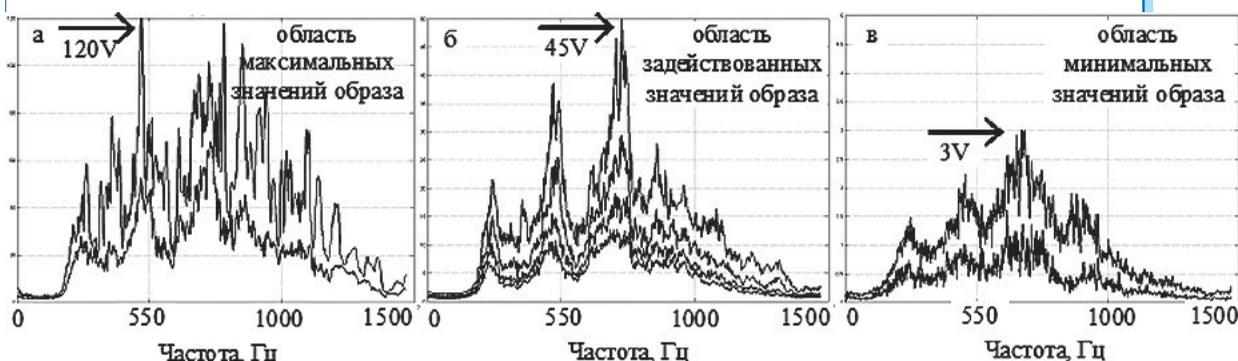


Рис. 5. Голосовой статистический образ диктора № 4. Область максимальных значений зависимостей «а»; область задействованных для идентификации значений «б»; область минимальных значений «в»

Максимальные значения порядковых статистик голосового статистического вектора представлены на рис. 5 «а» зависимостями квантилей экстремальных значений $\hat{x}_{1/n}, \hat{x}_{n/n}$ статистического вектора. Наблюдается значительная дисгармония крайних максимальных значений зависимостей, которая вызвана перемещением в эту область образа мощных спектральных составляющих. По оси абсцисс отложены значения частоты спектра сигнала, выведены значения до 1500 Гц, по оси ординат — амплитуда напряжения в вольтах.

Минимальные значения порядковых статистик голосового статистического образа представлены на рис. 5 «в» зависимостями квантилей экстремальных значений $\hat{x}_{1/n}, \hat{x}_{n/n}$ статистического вектора. В этой области наблюдается более спокойная картина крайних минимальных значений зависимостей, что объясняется перемещением в эту область образа спектральных составляющих низкой энергетикой. При этом наглядно прослеживается низкая «характерность» значений зависимостей квантилей.

При этом сохраняется значительная **стабильность** значений зависимостей квантилей при рассмотрении группы образов, принадлежащих одному диктору, т.е. при переходе от одного статистического вектора квантилей $[\hat{x}_{i/n}]$ к другому вектору квантилей одного и того же диктора, наблюдается значительная стабильность значений квантилей.

Диктор похож сам на себя. Для иллюстрации этого явления на рис. 6 представлены выборочные голосовые статистические образы дикторов № 1, 5, 7.

Наблюдается значительное внутрдикторное сходство образов по местоположению на оси частот максимальных значений зависимостей квантилей и по ширине полосы занимаемых ими частот, а также по величине энергетической составляющей зависимостей квантилей спектра, т.е. сохраняется значительная стабильность значений как частот, так и амплитуд. На рис. 6 по оси абсцисс отложены значения частоты спектра сигнала, выведены значения до 1500 Гц, по оси ординат — амплитуда напряжения в вольтах.

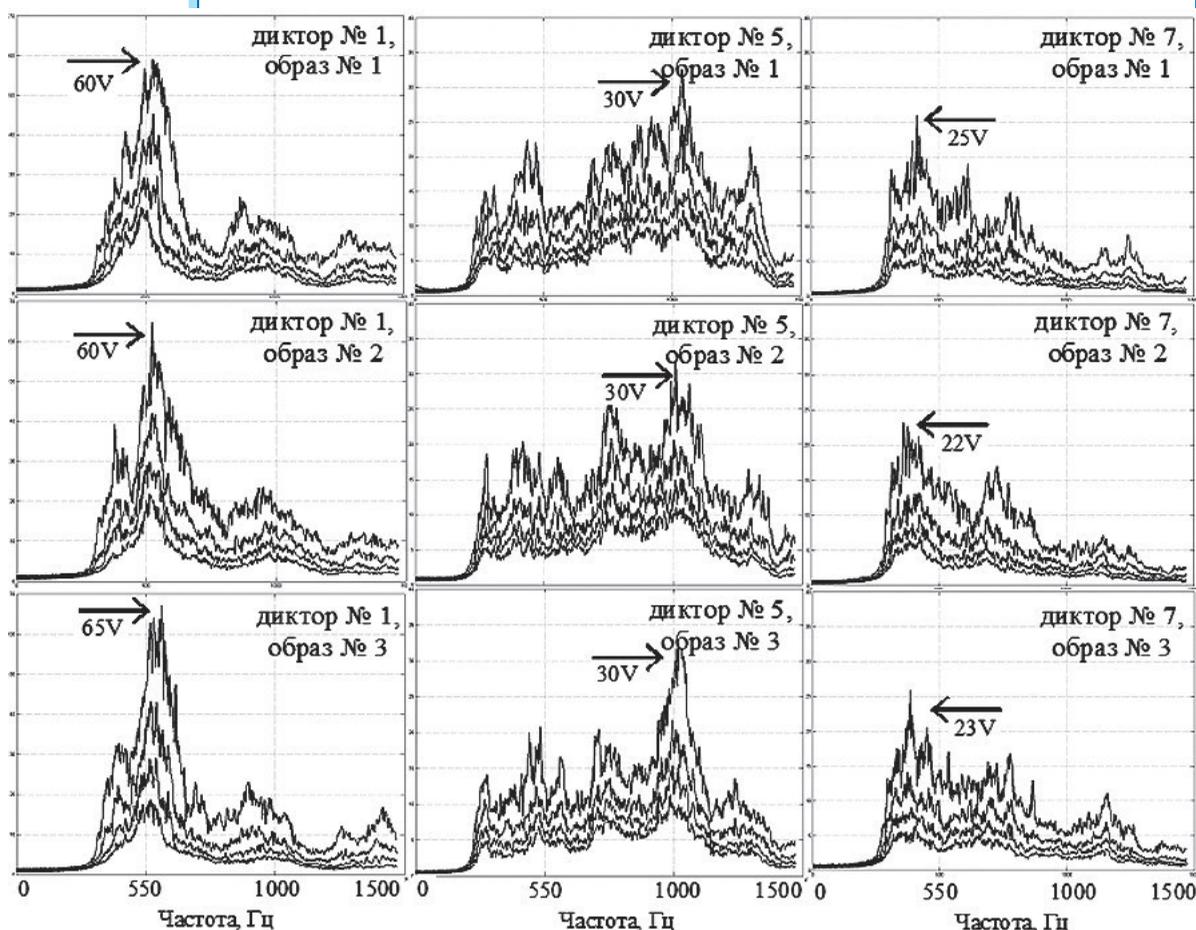


Рис. 6. Выборочные голосовые статистические образы дикторов

**4. Сопоставление векторов квантилей.
Эффективность разработанного метода идентификации**

Результатом формирования зависимостей квантилей $\hat{x}_\alpha : \hat{x}_{j/4}, j=1,2,3$ является голосовой статистический вектор квантилей постоянной размерности, который является объектом сопоставления. Поэтому, в зависимости от длительности идентифицируемых речевых сигналов, которые, как правило, отличаются, формируется различное количество векторов квантилей равной размерности. Для выполнения процедуры сопоставления значений выборочных квантилей статистических векторов используется одинаковое количество векторов для речевых сигналов базы данных и идентифицируемых сигналов. Подход к сопоставлению можно охарактеризовать как «каждый-с-каждым», при этом значения каждого выборочного квантиля каждого задействованного статистического вектора квантилей базы данных $Y_j, (j = 1, 2, 3, \dots, N)$ сопоставляются с соответствующими значениями каждого выборочного квантиля каждого статистического вектора квантилей идентифицируемого речевого сигнала $X_k, (k = 1, 2, 3, \dots, N)$ (рис. 7).

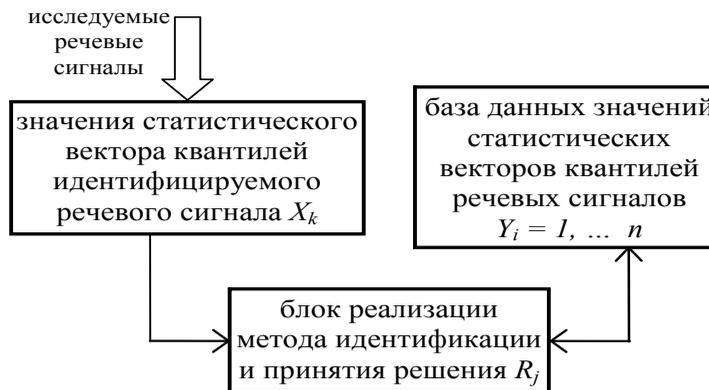


Рис. 7. Структурная схема реализации процедуры идентификации речевого сигнала

При этом вычисляются значения среднего абсолютного отклонения как наименее флюктуирующего параметра: $R_j = \text{mean}(\text{abs}(X_k - Y_i))$. Для каждой j -й процедуры сопоставления формируются два массива значений метрик R . Первый — «внутридикторные» метрики (массив «свой-свой») $R_{j,k=i} = \arg \min_i \|X_{j,k,n} - Y_{j,i,m}\|$, второй — «междудикторные» метрики (массив «свой-чужой»), рассчитанный по совокупности значений векторов квантилей всех «чужих» дикторов $R_{j,k \neq i} = \arg \min_i \|X_{j,k,n} - Y_{j,i,m}\|$, где j — номер квантиля, k , i — номера дикторов, n , m — номера статистических векторов квантилей.

По рассчитанным статистическим значениям массивов метрик $R_{j,k=i}$ (1275 значений метрик в примере) и $R_{j,k \neq i}$ (1225 значений метрик в примере) строим графики условных функций распределения (ФР) значений метрик для статистических векторов квантилей **идентичного источника** (массив «свой-свой») и статистических векторов квантилей **неидентичного источника** (массив «свой-чужой») (рис. 8 «а»). По функциям распределения значений метрик выведено решающее правило (рис. 8 «б»). Принятая пороговая величина вариативности Π_j для j -го признака является условной, с точки зрения допустимого уровня ошибок 1-го рода (пропуск, отбрасывание «своего» диктора) и 2-го рода (принятие «чужого» за «своего», ложная тревога).

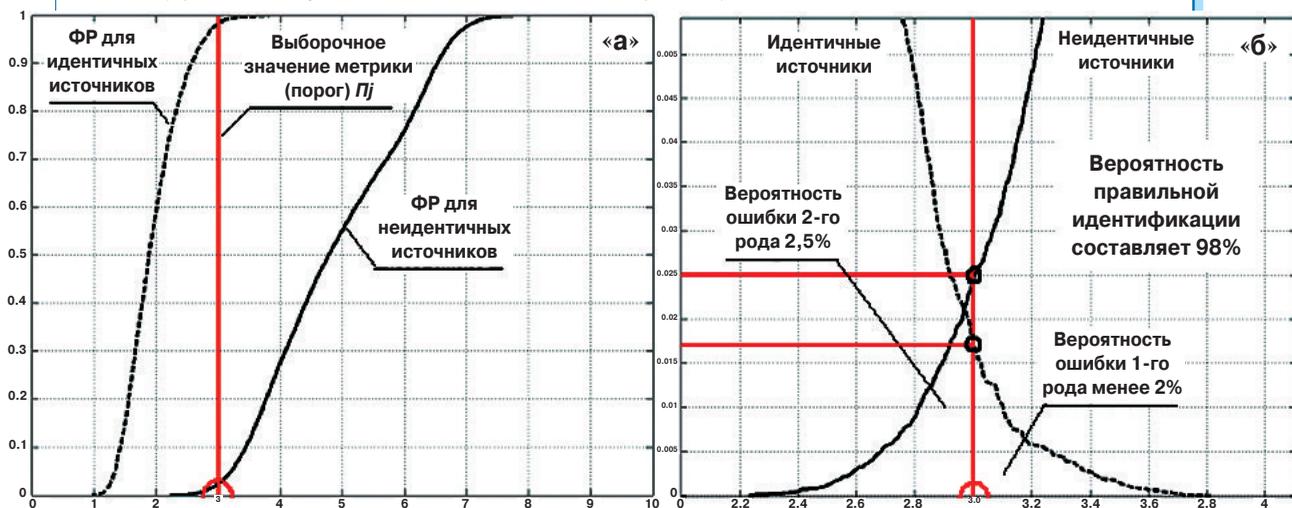


Рис. 8. Принятие решения об идентичности источников, решающее правило



Результаты экспериментальной проверки метода показали, что вероятность правильной идентификации исследуемых речевых сигналов составила **98%**, при вероятности ошибки 1-го рода **2%** и вероятность ошибки 2-го рода около **2,5%**. Допускается менять величину порога с целью изменения условия принятия решения об идентичности источников. Уменьшая значение вероятности пропуска, получим возрастание количества ложных тревог, и наоборот, уменьшая вероятность возникновения ложной тревоги, увеличивать вероятность пропуска.

Выводы

По результатам проведения экспериментов можно сделать выводы, что разработанный новый метод идентификации речевых сигналов [7], основанный на применении зависимостей от частоты спектра сигнала квантилей вариационных рядов — порядковых статистик спектра Фурье речевого сигнала, обладает следующими параметрами:

1. В соответствии с алгоритмом метода, для идентификации речевого сигнала, **не требуется** проводить следующие исследования: влияние на речевой сигнал психофизического состояния диктора, смыслового наполнения речи, акустических и лингвистических признаков речи.
2. В силу нормализации распределения значений спектра в составе выборочных квантилей, достигается повышение эффективности идентификации речевых сигналов известными методами статистической проверки гипотез по критерию Неймана-Пирсона. Метод идентификации приобретает свойства устойчивости к изменениям как статистических характеристик речи, так и статистических характеристик каналов передачи речевых данных.
3. Результаты идентификации слабо зависят от разницы в уровне вокализации речевых сигналов идентифицируемого источника и образцовых речевых сигналов базы данных.
4. Для идентификации требуется речевой сигнал минимальной длительности 30–40 секунд (без пауз), применяется мажоритарная схема сопоставления.
5. Результаты идентификации не зависят от языка речи.
6. Значения статистического вектора квантилей речевого сигнала **не усредняются**, следовательно, не происходит утрат информативного, характеризующего диктора наполнения сигнала. Реализована детализация статистического вектора.
7. Реализована возможность перехода от сопоставления параметров многомерных суммарных **ПРВ** к сопоставлению числовых значений **векторов** квантилей вариационного ряда.

Комплексное использование перечисленных отличительных особенностей и связанных с ними положительных эффектов позволяет обеспечить повышение эффективности существующих и перспективных систем спектральной идентификации речевого сигнала.



Кузнецов М.В.

Метод идентификации речевых сигналов на основе порядковых статистик амплитудных спектров

Литература

1. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи / Под. ред. О.И. Шелухина. М.: Радио и связь, 2000. С. 98–106.
2. Акишин Б.А., Хвастунов Ю.А. Цифровой спектральный анализ речевых сигналов: Учебное пособие. Казань: Изд-во Казан. гос. техн. ун-та, 2000. 48 с.
3. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ./ Под ред. М.В. Назарова и Ю.Н. Прохорова. М.: Радио и связь, 1981. 496 с.
4. Автоматическое распознавание и синтез речи. Сборник научных трудов НАН Беларуси, Институт технической кибернетики / Научный редактор д.т.н. Лобанов Б.М. Минск, 2000. 172 с.
5. Женило В.Р., Минаев В.А. Компьютерные технологии в криминалистических фоноскопических исследованиях и экспертизах: Учебное пособие. М.: Академия МВД РФ, 1994. 139 с.
6. Дейвуд Дж. Порядковые статистики: Пер. с англ. М.: Мир, 1989. 540 с.
7. Патент України на винахід № 80587 МПК (2006). Спосіб спектральної ідентифікації мовного сигналу / М.В. Кузнецов, В.Л. Селетков (Україна). Опубліковано бюл. «Промислова власність» № 16 10.10.2007 р.

Сведения об авторе

Кузнецов Максим Владимирович —

автор пяти научных публикаций, четырёх патентов на изобретения; область научных интересов: идентификация сложных сигналов. Ttmarun@rambler.ru