АДАПТИВНОЕ УПРАВЛЕНИЕ И МЕТОДЫ ИДЕНТИФИКАЦИИ

УДК 621.317+681.849

В.И. Соловьев, О.В. Рыбальский, В.В. Журавель, А.Н. Шабля, Е.В. Тимко

УЧЕТ МНОГОФАКТОРНОСТИ ХАРАКТЕРИСТИК ГОЛОСА В ЗАДАЧАХ ИДЕНТИФИКАЦИИ ДИКТОРА

Ключевые слова: вероятность, временное окно, гласный звук, диктор, идентификация, кривые ошибок, спектр, точка пересечения, фонограмма, экспертиза, эффективность.

Введение

При тестировании на специализированных базах данных наиболее совершенных систем идентификации дикторов их минимальная эффективность, оцениваемая величиной вероятности ошибки в точке пересечения кривых ошибок первого и второго рода, составляет всего несколько процентов [1–6]. Однако часто тестирование одних и тех же систем на различных тестовых базах данных дает весьма отличающиеся друг от друга результаты [4–6]. Это поясняется доминированием в разных тестовых базах данных факторов, оказывающих особое влияние на результаты идентификации диктора по характеристикам голоса.

На вариативность характеристик голоса диктора, как правило, влияют язык и диалекты, контекст речи, эмоциональное состояние диктора, длительность фонограммы и др. [1–6]. Поэтому апробация и оценка эффективности таких систем с охватом всех основных факторов практически невозможна. Основная сложность состоит в количественной формализации ряда конкретных, влияющих на характеристики голоса диктора факторов, каждый из которых влияет на положение точки пересечения кривых вероятностей ошибок первого и второго рода, являющейся, как правило, мерой эффективности таких систем.

Поэтому многие практикующие эксперты, несмотря на высокую эффективность ряда систем, весьма скептически относятся к возможности автоматической идентификации диктора [4–6].

Цель работы — создание метода идентификации диктора, учитывающего множество основных факторов, влияющих на параметры характеристик голоса.

Полагаем, что наиболее удобная реализация данного метода состоит в одновременном учете влияния двух различных факторов на зависящий от них параметр, являющийся определяющим при решении задачи идентификации диктора. При парном подборе (или удачном отборе) таких факторов можно обеспечить принципиальную возможность косвенного учета их практически неограниченного количества. В общем виде такая постановка задачи может быть представлена соотношением

$$P_{\Sigma}(K_C, l) = \sum_{i=1}^{N} \sum_{l=1}^{L_C} K_C(K_C, l), \qquad (1)$$

где P_Σ — параметр, зависящий от влияния двух факторов, K_C — первый из влияющих факторов, действующий на интервале C, l — второй из влияющих факторов, действующий на интервале C.

Такой метод реализован в разработанной системе идентификации диктора «Аватар».

Многофакторность характеристик голоса

Учет влияния многофакторности на характеристики голоса приводит не к графикам ошибок первого и второго рода в двумерном пространстве, а к N-мерным поверхностям ошибок первого и второго рода в N-мерном пространстве.

Рассмотрим проекцию N-мерного пространства поверхностей ошибок первого и второго рода на трехмерное пространство (рис. 1, 3D-поверхности ошибок первого и второго рода рассчитаны и нормированы для двух факторов — близости спектральных характеристик исследуемых звукосочетаний и длительности фонограммы). Далее все графики и примеры реализованы на модулях системы идентификации и верификации дикторов «Аватар» [7–10].

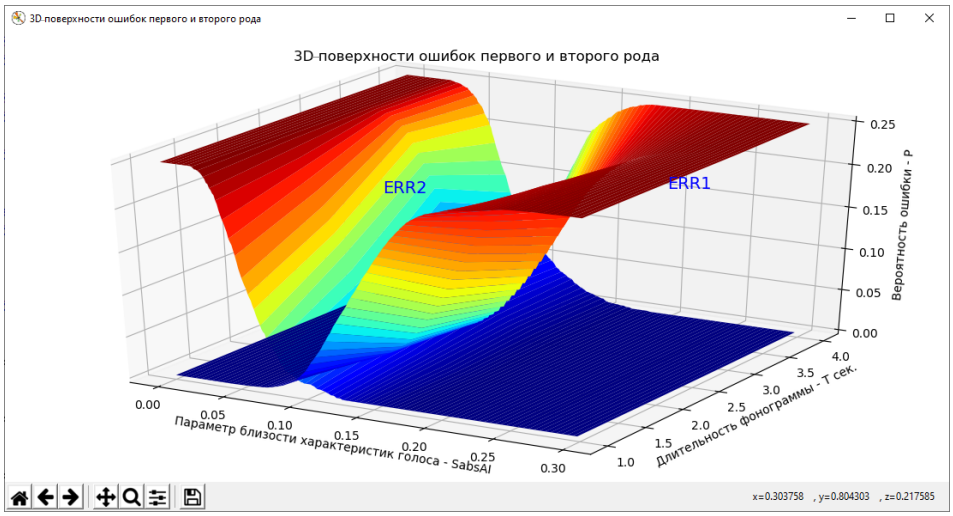

Рис. 1

Данная поверхность построена на основе большой базы данных, полученных для одного и разных дикторов при сравнении характеристик их голосов для звукосочетания [А][И] на фонограммах малой (до 5 с) длительности. Язык фонограмм — украинский. По оси OX представлен фактор близости характеристик голосов в системе «Аватар» (SabsAI), который находится как сумма модулей абсолютной разности суммарных спектров для этих звуков на каждом отсчете. По оси OY — длительность фонограмм T в секундах. По оси OZ — вероятности ошибок первого и второго рода, нормированные для этих двух факторов (особенности нормирования вероятностей ошибок будут рассмотрены ниже). Естественно, что точки пересечения поверхностей ошибок лежат на пространственной линии и зависят от двух принятых параметров. Важными здесь являются величины вероятностей в точках пересечения. Эти величины гораздо меньше, чем для двумерных графиков ошибок, что естественно ввиду учета двух факторов. В многомерном пространстве факторов (при построении и расчете многомерных поверхностей ошибок) эти величины будут уменьшаться при росте количества учитываемых факторов.

Для иллюстрации этого рассмотрим тот же график, но нормированный по вероятностям ошибок только по одной координате (OX) (рис. 2, 3D-поверхности

ошибок первого и второго рода рассчитаны и нормированы для одного фактора — близости спектральных характеристик). На этом графике вероятности в точках пересечения трехмерных поверхностей находятся «выше» по шкале величины ошибок в несколько раз. Для более полной иллюстрации на рис. 3 приведен двумерный срез поверхностей ошибок первого и второго рода, который соответствует обычному классическому представлению графиков ошибок.

Подобные трехмерные поверхности можно построить практически для любых пар параметров характеристик голоса (при наличии соответствующих программных инструментов).

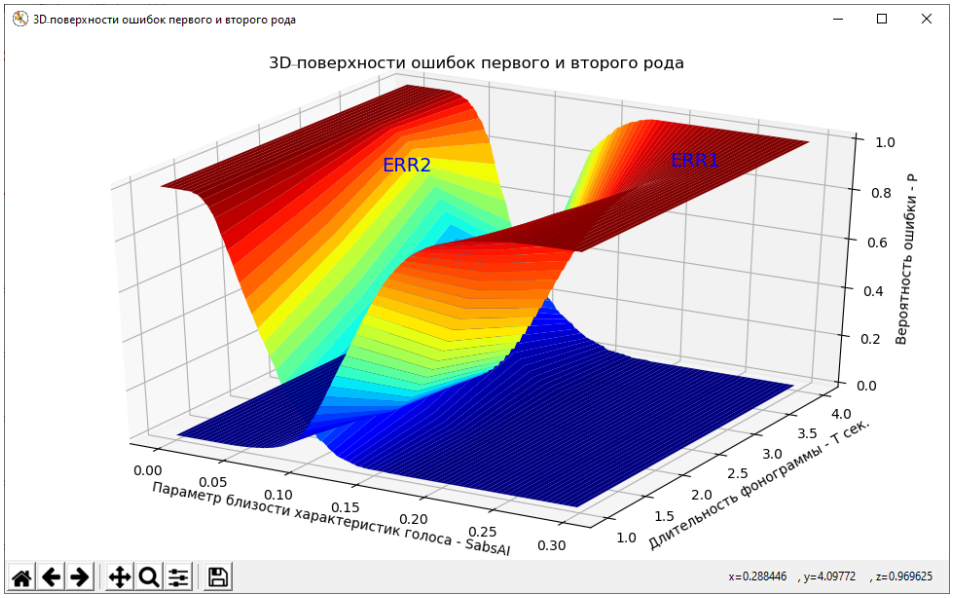

Рис. 2

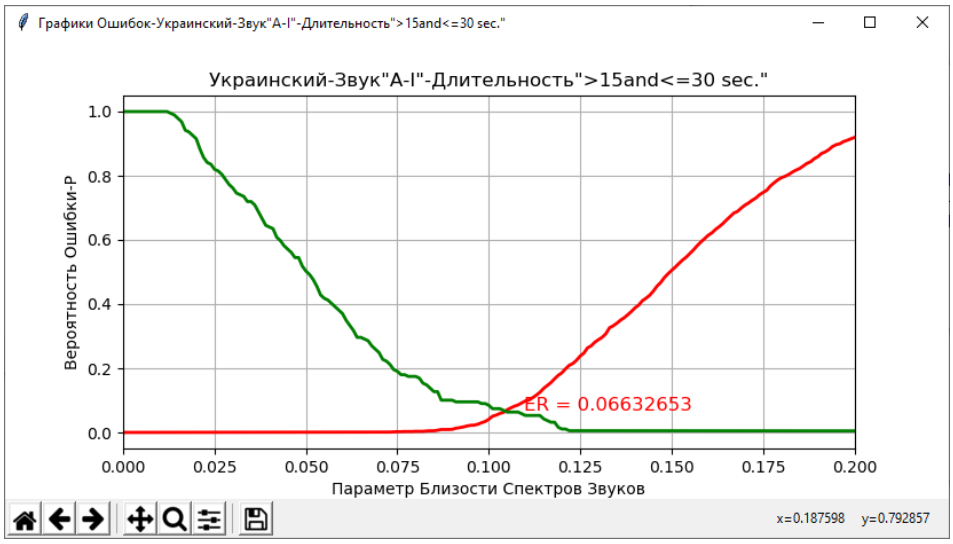

Рис. 3

Очевидно, что при наличии в тестовой базе данных большого количества факторов, прямо или косвенно влияющих на характеристики голоса, точность оценки эффективности систем при увеличении количества данных будет расти. При этом если конкретная система идентификации также учитывает эти факторы, то эффективность системы при тестовых испытаниях будет повышаться (снижа-

ется вероятность ошибок в точке пересечения графиков ошибок). Это касается любых факторов, кроме длительности фонограмм. Но, как правило, это один из самых существенных факторов при идентификации диктора, поскольку для фонограмм малой длительности фактически устанавливает определенный предел эффективности любой системы.

Реализация метода, учитывающего многофакторность характеристик голоса диктора

Классический метод проведения исследований и разработок в области систем идентификации дикторов обычно подразумевает определенную последовательность. На основе известных представлений о факторах и параметрах, влияющих на характеристики голоса диктора, выделяются и количественно описываются эти параметры из большого статистического материала. Далее определяются количественные взаимосвязи между характеристиками голоса и выделенными параметрами. Такой метод практически не зависит ни от методологии исследований, ни от технологий построения систем. Эта схема используется и при разработках на основе технологий нейронных сетей. При таком методе исследователи и разработчики на основе известных представлений и интуиции выделяют основные параметры, существенно влияющие на характеристики голоса диктора. Применение нейронных сетей, по существу, ничего не изменяет при таком подходе, поскольку если нейронная сеть и способна выделить некоторые новые закономерности, то это осуществляется в рамках определенного перечня параметров, подаваемых на вход нейронной сети для обучения. Самый главный недостаток этого метода — необходимость выбрать из субъективных предпосылок определенный перечень параметров для разработки и исследований, который, в конечном счете, и определяет эффективность разрабатываемой системы. Попытки применения всей полноты информации, содержащейся, например, в звуковой волне с речевыми фрагментами, до настоящего времени не известны. Кроме того, несмотря на тенденции применения на входе нейронных сетей всей совокупности исходных данных, примеры эффективных решений, полученных для этого технологического направления, до настоящего времени отсутствовали [4–6].

Для разрешения возникшего противоречия рассмотрим еще один метод, который назовем методом «атомарных» структур. Согласно этому методу из речевых сигналов выделяются структуры, зависящие от совокупности основных факторов, влияющих на процесс идентификации диктора. При данном методе все существенные факторы, влияющие на характеристики голоса, будут косвенно учитываться на уровне этих структур. Последующие решения по идентификации на основе всей фонограммы будут приниматься по комбинаторной совокупности огромного числа этих «атомарных» структур.

Под «атомарными» структурами речи понимаем спектры любых фрагментов гласных звуков, выделяемых во временном окне длительностью 20 мс.

Рассмотрим дискретное неортогональное время-частотное преобразование сигнала звукового диапазона частот во временном окне длительностью 20 мс. В качестве базиса воспользуемся вейвлетом Морле [11]

$$C_{mor}(t) = \frac{e^{j2\pi F_C(t)} e^{\frac{-t^2}{F_b}}}{\sqrt{\pi F_b}}, \qquad (2)$$

где t — время, F_b — параметр ширины вейвлета, F_C — центральная частота (частота гетеродинирования) вейвлета при сканировании сигнала в окне длительностью 20 мс [11].

На основе вейвлета Морле реализуем спектральное преобразование

$$Y_{FC} = \sum_{i=1}^{N} A_i(t_i) \times C_{mor}(t_i, F_C)\ ,\ i = 1, 2, \ldots, N, \tag{3}$$

$$S_{FC} = \sqrt{(|Y_{FC}|)\frac{|Y_{FC}|}{N}}\ , \tag{4}$$

где $A_i(t_i)$ — дискретные отсчеты звукового сигнала во временном окне длительностью 20 мс, Y_{FC} — результат комплексного преобразования сигнала в частотную область, F_C — дискретные значения частот с интервалом сканирования по частоте D_{FC} = 1Гц, S_{FC} — нормированные уровни спектральных компонент, N — количество усреднений на каждый отсчет, t_i — дискретный i-й временной отсчет.

При этом рассмотрим избыточные преобразования, в которых число отсчетов во временной области меньше числа отсчетов в частотной области. Так, для примера, возьмем произвольный фрагмент длительностью 20 мс речевого сигнала звука [А] с частотой дискретизации 44100 Гц. Тогда число дискретных отсчетов на участке длительностью 20 мс составляет N = 882. Для представляемого неортогонального преобразования на основе вейвлета Морле с шагом сканирования в частотной области D_{FC} = 1 Гц общее максимально возможное число шагов сканирования в частотной области в выбранном диапазоне — 22050 отсчетов [11].

Уравнение (3) чисто формально является ковариационной функцией между дискретными амплитудами звукового сигнала и гармоническими функциями во временном окне. Уравнение (4) может трактоваться как обычный спектр, но с шагом по частоте в 1Гц. В точках F_C, кратных 50 Гц, результаты этого преобразования совпадают с результатами быстрого преобразования Фурье (БПФ) (с точностью до функции Гаусса, используемой для сглаживания влияния малых размеров временного окна на спектр).

Особенностью данного метода является точность определения положения локальных максимумов спектра в малом временном окне, составляющая примерно 1 Гц. Но именно положением локальных максимумов спектра и определяется с физической точки зрения точность оценки частоты основного тона и практически всех функций спектра, используемых в экспертизе.

Количество в частотной области частот градаций спектра, используемых при данном методе, с точки зрения классических представлений является избыточным. Однако эта избыточность позволяет значительно повысить точность оценок для любой методологии проведения исследований и разработок и, в частности, при решении задач на основе нейронных сетей глубокого обучения [7–10].

Далее будем рассматривать спектры гласных звуков на временном интервале в 20 мс. Известно, что спектры практически любого гласного звука для одного диктора весьма вариабельны, как по виду спектра, так и по положениям максимумов в динамике, даже при произнесении одного гласного звука [7–10]. Рассмотрим модель, характеризующую динамику спектра локально произносимого гласного звука.

Выделим произвольный гласный звук в фонограмме определенного диктора. В исследованиях это выделение осуществлялось автоматическим программным модулем в специальном звуковом редакторе (данный модуль встроен в систему «Аватар»). Этот редактор содержит специальный программный модуль на основе нейронных сетей глубокого обучения и обеспечивает автоматическое выделение гласных звуков, независимо от языка, контекста речи и диктора. В стандарте международной транскрипции это шесть гласных звуков: [A], [E], [I], [I:], [O], [U]. Выбор множества гласных звуков — выбор некоторого усредненного множества

гласных звуков для различных языковых групп. Этот выбор не опирается на конкретное лингвистическое описание гласных фонем, звуков и их сочетаний для конкретных языковых групп. В своей основе это — гласные звуки индоевропейских языков.

На основе изложенного метода был сформирован Dataset для обучения нейронной сети идентификации диктора на уровне «атомарных» спектров гласных звуков длительностью 20 мс. Этот метод содержал миллионы фрагментов спектров для разных дикторов, в том числе сотни тысяч фрагментов для одного и того же диктора.

Необходимо отметить, что в системе с модулем фонемической машины применяется технология идентификации звуков, которая отличается интерпретацией звуков от классических представлений. В частности, для эффективного решения задач идентификации дикторов не выделяются звуки, физиологически воспринимаемые органами слуха. В разработанной фонемической машине модели звуков речи представляются в виде определенной совокупности нескольких «атомарных» составляющих для каждого звука. Множество «атомарных» составляющих для каждого звука может частично пересекаться со множеством «атомарных» составляющих для других звуков. Так, например, множество составляющих структур звука [А] пересекается со множеством структур звука [О] (могут быть и другие пересечения). Аналогично звук [И] — со звуком [Ы]. При усредненном представлении в виде спектров эти структуры можно интерпретировать как усредненные спектры соответствующих звуков конкретного диктора. Но эти усредненные спектры «атомарных» звуков не полностью соответствуют физиологически воспринимаемым звукам. Таким образом, в применяемой модели фонемической машины выделяются структурные составляющие звуков речи.

При обучении решалась задача бинарной классификации диктора (ОН–НЕ ОН) по близости фрагментов «атомарных» спектров гласных звуков.

Важным фактором обучения нейронной сети при бинарной идентификации дикторов (с точки зрения близости характеристик их голосов) является множественность моделей. Поэтому разрабатывались модели идентификации диктора по близости спектральных характеристик как для каждого из шести гласных звуков, так и их сочетаний. В частности, использовались следующие их сочетания: [A][E], [A][I], [A][I:], [A][O], [A][U], [E][I], [E][I:], [E][O], [E][U], [I][I:], [I][O], [I][U], [I:][O], [I:][U], [O][U]. Для каждого из сочетаний «атомарных» спектров в процессе обучения были получены отдельные модели. Входом нейронной сети при обучении были «атомарные» спектры гласных звуков и их сочетания. Выходом — вероятность того, что два «атомарных» спектра принадлежат одному диктору.

Иллюстрация обучения нейронной сети для бинарной классификации звукосочетания [А] [Е] приведена на рис. 4.

Эффективность идентификации диктора на уровне «атомарных» структур спектров относительно низка. Однако в данном методе идентификация диктора по совокупности «атомарных» структур спектров в двух фонограммах осуществляется по огромному количеству комбинаторных сочетаний «атомарных» спектров гласных звуков. Так, например, в среднем для фонограммы длительностью 10 с общая длительность гласных звуков составляет не менее 3–4 с (с учетом пауз речи и согласных звуков). При рассмотрении 20 мс «атомарных» интервалов их будет около 200. Существенным фактором количества «атомарных» структур в представляемом исследовании является сканирование гласных звуков во временном окне длительностью 20 мс. Длительность гласного звука при среднем темпе речи — 30–60 мс [12, 13].

Как показывает анализ динамики изменения спектров одного гласного звука, «атомарные» спектры звука на интервале в 20 мс весьма вариабельны для одного звука, как по положению максимума спектра, так и по структуре спектра. В данном случае в исследованиях и разработке применялось сканирование гласного звука в окне длительностью 20 мс с интервалом сканирования не более 5 мс.

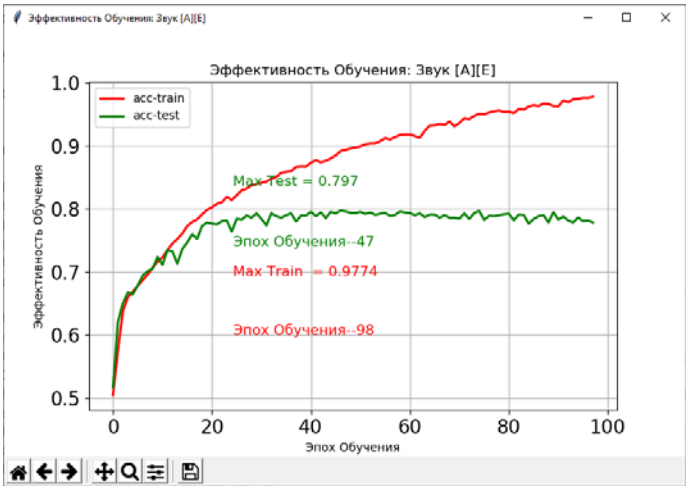

Рис. 4

Таким образом, для фонограммы длительностью 10 с число различных «атомарных» спектров может достигать 1000. А число различных комбинаторных сочетаний «атомарных» спектров в двух фонограммах длительностью 10 с — порядка $N = 1000000$. Это огромная статистика по различным значениям вероятности P идентификации диктора на уровне «атомарных» спектров гласных звуков (выход нейронной сети при бинарной классификации — вероятность P «ОН–НЕ ОН» идентификации диктора). Далее принятие решения по всей комбинаторной совокупности «атомарных» спектров гласных звуков осуществляется на основе классических представлений теории вероятностей и математической статистики. Осуществляется расчет распределения вероятности P по всей совокупности величин. Определяется математическое ожидание (P_{sr}) и дисперсия распределения (D). При $P_{sr} > 0{,}5$ принимается решение «ОН», при $P_{sr} < 0{,}5$ — «НЕ ОН». Ошибка принятия решения определяется по точности вычисления P_{sr} и является функцией числа комбинаторных сочетаний N и вида функции распределения (для распределения Гаусса достаточно N и D).

Предложенный метод на основе «атомарных» спектров гласных звуков не требует выбора отдельных конкретных факторов и параметров характеристик голоса при разработке системы идентификации диктора. Это поясняется тем, что практически все факторы, прямо или косвенно влияющие на параметры характеристик голоса диктора, учитываются в спектрах «атомарных» структур.

Результаты экспериментов показывают весьма высокую эффективность данного метода при анализе фонограмм малой длительности, в частности фонограмм с длительностью речи менее 1с (при наличии в них гласных звуков). На рис. 5 показана идентификация диктора системой, основанной на предложенном методе, по фонограммам длительностью менее 1с, а также идентификация диктора на основе метода «атомарных» структур по звукам [А] из двух различных фонограмм для одного и того же диктора. Длительность звуков составляет 63 и 81 мс.

С точки зрения экспертизы речевых фонограмм подобная идентификация (по столь коротким фрагментам) является «экзотикой». Но в рамках изложенной методологии это реальная возможность, реализованная в действующей системе идентификации.

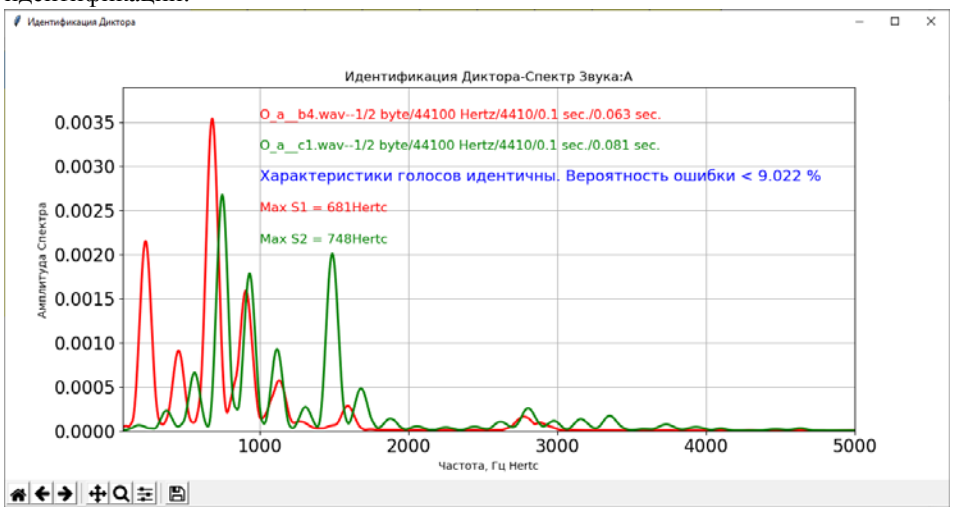

Рис. 5

При этом, как показывают результаты различных тестирований, даже для одиночных звуков различных фонограмм вовсе не обязательно, чтобы ошибка принятия решения была весьма высокой. Так, например, на рис. 6 представлены результаты для звуков [А] одного и того же диктора из записей на различных фонограммах. Но звуки вырезаны из весьма близкого контекста одного и того же диктора, поэтому ошибка принятия решения, с учетом близости множества существенных факторов, определяющих характеристики голоса, может быть весьма мала.

Необходимо отметить очень важный момент проведенных исследований и разработок. Увеличение числа дикторов, вариабельности речи, количества языковых групп для базы данных Dataset, начиная с некоторого момента, практически не оказывает влияния на «атомарные» модели (как по массивам обучения, так и тестирования). По-видимому, на данном этапе исследований и разработок можно предположить, что представляемый метод имеет весьма высокую степень обобщения на различные языковые группы. Он не зависит от диктора, контекста и многообразия различных факторов и параметров речи.

Рис. 6

Заключение

Предложен метод, обеспечивающий рациональный учет многофакторности влияния различных параметров на характеристики голоса при идентификации диктора. В основе метода лежит выделение из голосовых сигналов гласных звуков «атомарных» структур во временном окне длительностью 20 мс. На спектры этих структур влияют все основные факторы, характеризующие индивидуальность голоса конкретного диктора. Решение об идентичности голосов дикторов, записанных на разных фонограммах, осуществляется на основе комбинаторики «атомарных» спектров гласных звуков в обеих фонограммах. Метод показал высокую эффективность при экспертизе фонограмм малой длительности.

В.І. Соловйов, О.В. Рибальський, В.В. Журавель, О.М. Шабля, Є.В. Тимко

УРАХУВАННЯ БОГАТОФАКТОРНОСТІ ХАРАКТЕРИСТИК ГОЛОСУ В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ ДИКТОРА

При тестуванні на спеціалізованих базах даних найбільш досконалих систем ідентифікації диктора їх мінімальна ефективність, що оцінюється величиною ймовірності помилки в точці перетину кривих помилок, становить лише кілька відсотків. Однак відомо безліч факторів, що впливають на варіативність характеристик голосу диктора, кожний з яких має свій, відмінний від інших, вплив на результати ідентифікації диктора за характеристиками голосу. Складність створення і тестування систем ідентифікації диктора полягає в необхідності кількісної формалізації ряду конкретних факторів, що впливають на характеристики його голосу. Розглянуто запропонований метод урахування безлічі чинників, які впливають на параметри характеристик голосу диктора, що забезпечує принципову можливість непрямого урахування їх практично необмеженої кількості. Відповідно до цього методу з мовних сигналів виділяються «атомарні» структури, які залежать від сукупності основних факторів, що впливають на процес ідентифікації диктора. За таким методом всі істотні фактори, що впливають на характеристики голосу, будуть побічно враховуватися на рівні цих структур. Експертні рішення приймаються за комбінаторною сукупністю величезного числа цих «атомарних» структур. Під «атомарними» структурами мовлення розуміються спектри будь-яких фрагментів голосних звуків, які виділяються в часовому вікні тривалістю 20 мс. «Атомарні» структури виділяються в автоматичному режимі. Запропонований метод забезпечує раціональне урахування богатофакторності впливу різних параметрів, оскільки на спектри цих структур впливають всі основні фактори, що характеризують індивідуальність голосу конкретного диктора. Рішення щодо ідентичності голосів дикторів, записаних на різних фонограмах, здійснюється на основі комбінаторики «атомарних» спектрів голосних звуків в обох фонограмах. Метод показав високу ефективність при експертизі фонограм малої тривалості.

Ключові слова: ймовірність, часове вікно, голосний звук, диктор, ідентифікація, криві помилок, спектр, точка перетину, фонограма, експертиза, ефективність.

V.I. Solovyov, O.V. Rybalsky, V.V. Zhuravel, A.N. Shablya, E.V. Tymko

TAKING INTO ACCOUNT THE MULTIFACTORIAL CHARACTER OF VOICE CHARACTERISTICS IN THE PROBLEMS OF SPEAKER IDENTIFICATION

When testing the most advanced speaker identification systems on specialized databases, their minimum efficiency, estimated by the error probability at the point of intersection of the error curves, is only a few percent. However, many factors are known that affect the variability of the characteristics of the speaker's voice, each of

which has its own, different from the others, influence on the results of the speaker's identification by the characteristics of the voice. The complexity of creating and testing speaker identification systems is the need to quantitatively formalize a number of specific factors that affect the characteristics of his voice. The article discusses the proposed method for accounting for a variety of factors affecting the parameters of the characteristics of the speaker's voice, which provides the fundamental possibility of indirectly accounting for their practically unlimited number. According to this method, «atomic» structures are distinguished from speech signals, which depend on the totality of the main factors that affect the speaker's identification process. With this method, all significant factors affecting the characteristics of the voice will be indirectly taken into account at the level of these structures. Subsequent decisions are made on the combinatorial set of a huge number of these «atomic» structures. «Atomic» speech structures are understood as the spectra of any fragments of any vowel sounds allocated in a time window of 20 ms. «Atomic» structures are selected automatically. The proposed method provides a rational consideration of the multifactorial influence of various parameters, since the spectra of these structures are influenced by all the main factors that characterize the individuality of the voice of a particular speaker. The decision on the identity of the voices of the announcers recorded on different phonograms is carried out on the basis of combinatorics of «atomic» spectra of vowel sounds in both phonograms. The method has shown high efficiency in the examination of phonograms of short duration.

Keywords: probability, time window, vowel sound, speaker, identification, error curves, spectrum, intersection point, phonogram, expertise, efficiency.

1. Lei Y., Scheffer N., Ferrer L., McLaren M. A novel scheme for speaker recognition using a phonetically-aware deep neural network. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2014. P. 1695–1699.
2. Deep neural networks for extracting baum-welch statistics for speaker recognition. P. Kenny, V. Gupta, T. Stafylakis, P. Ouellet, J. Alam. Odyssey : The Speaker and Language Recognition Workshop. 2014. P. 293–298.
3. Kassin S.M., Dror I.E., Kukucka J. The forensic confirmation bias: problems, perspectives and proposed solutions. *Journal of Applied Research in Memory and Cognition*. 2013. 2(1). P. 42–52.
4. Satyanand Sing. Forensic and automatic speaker recognition system. *International Journal of Electrical and Computer Engineering (IJECE)*. 2018. **8**, N 5. P. 2804–2811.
5. Amali Mary Bastina, Rama N. Biometric identification and authentication providence using fingerprint for cloud data access. *International Journal of Electrical and Computer Engineering*. 2017. 7(1). P. 408–416.
6. John H.L. Hansen, Taufiq Hasan. Speaker recognition by machines and humans. *IEEE Signal Process. Mag.* 2015. **32**, N 6. P. 74–99.
7. Solovyov V.I., Rybalskiy O.V., Zhuravel V.V., Semenova N.V.Analyzing the models of speech recognition on the basis of neural networks of deep learning for examination of digital phonograms. *Cybernetics and Systems Analysis*. 2021, **57**, N 1. P. 133–138. https://doi.org/10.1007/s10559-021-00336-y.
8. Solovyov V.I., Rybalskiy O.V., Zhuravel V.V., ZheleznyakV.K. Application of neuron networks of deep learning for exposures editing of digital phonograms. *Proceedings of the National Academy of Sciences of Belarus. Physical-technical series*. 2020. **65**, N 3. P. 383–389. https://doi.org/10.29235/1561-8358-2020-65-3-383-389.
9. Solovyov V.I., Rybalskiy O.V., Zhuravel V.V. Method of exposure of signs of the digital editing in phonograms with the use of neuron networks of the deep learning. *Journal of Automation and Information Sciences*. 2020. **52**, N 1. P. 22–28. https://doi.org//10.1615/JAutomatInfScien.v52.i1.30.
10. Solovyov V.I., Rybalskiy O.V., Zhuravel V.V. Verification of fundamental fitness of neuron networks of the deep educating for the construction of the system of exposure of editing of digital phonograms. *Cybernetics and Systems Analysis*. 2020. 56, N 2. P. 326–330. https://doi.org//10.1007/s10559-020-00249-2.
11. Mallat S. A wevlet tour of signal processing. New York : AcademicPress, 1999, 671 p.
12. Фланаган Джеймс Л. Анализ, синтез и восприятие речи. Пер. с англ. А. А. Пирогова. М. : Связь, 1968. 397 с.
13. Фант Г. Акустическая теория речеобразования. Перевод с английского Л.А. Варшавского и В.И. Медведева. Под редакцией В.С. Григорьева. М. : Наука, 1964. 284 с.

Получено 05.07.2021
После доработки 10.08.2021