

АЛГОРИТМ ПОСТРОЕНИЯ СТАТИСТИЧЕСКОГО ДИСКРЕТНО-КОНТИНУАЛЬНОГО ОПИСАНИЯ ДЛИТЕЛЬНОСТИ ЗВУКОВ ПОТОКА ОСМЫСЛЕННОЙ РЕЧИ ДИКТОРА¹

Аннотация.

Актуальность и цели. Основными проблемами при разработке алгоритмов и программ, реализующих аутентификацию по голосу, являются следующие: вариации голоса пользователя (голос способен изменяться в зависимости от состояния здоровья, возраста, настроения и т.д.); наличие шумовой компоненты. Решение этих проблем позволит применять голосовую технологию аутентификации, которая обеспечит наилучшую защиту персональных данных, простоту в применении, и которая является наиболее дешевой среди существующих технологий идентификации личности.

Материалы и методы. В работе использованы численные и цифровые методы обработки сигналов, спектральные методы, методы математической статистики и временных рядов, а также искусственного интеллекта и распознавания образов. В основу построения фрагментатора положена непрерывно-дискретная модель обработки речи, которая в сочетании с узкополосным фильтром позволяет определять среднюю длительность звука.

Результаты. Показано, что качественный классификатор речи тон/шум должен давать выходные данные «0» и «1», длительность которых описывается непрерывно-дискретным распределением значений длительности интервалов между участками тональных звуков, распределенных по нормальным законам. Дискретная часть распределения образуется дискретным характером появления в речи тональных и шумовых звуков, а также их сочетаний (пар, троек, четверок и т.д.). Непрерывная (непрерывная) часть распределения значений длин звуков обусловлена нестабильностью речи при смене темпа произношения. Приведен метод вычисления средней длины одного звука осмысленной речи. Данное исследование позволило построить автомат по определению средней длины звука на различных участках звукового сигнала.

Выводы. Предложен численный алгоритм идентификации речи отдельного диктора, позволяющий производить синхронизацию участков речи. Использование разработанного алгоритма позволило уточнить значения параметров, характеризующих статистическое описание длительности интервалов между шумовыми звуками речи и между тональными звуками. Проведенное исследование позволило построить автомат по определению средней длины звука на различных участках звукового сигнала. Полученные результаты являются базой для построения нейросетевых технологий аутентификации.

Ключевые слова: численные методы, цифровая обработка сигналов, биометрические системы, голосовая аутентификация.

I. V. Boykov, A. I. Ivanov, D. M. Kalashnikov

AN ALGORITHM FOR CONSTRUCTING A STATISTICAL DISCRETE-CONTINUUM DESCRIPTION

¹ Работа выполнена при финансовой поддержке РФФИ (грант 16-01-00594).

OF SOUND DURATION OF ANNOUNCER'S MEANINGFUL SPEECH FLOW

Abstract.

Background. Main problems in development of algorithms and software for implementing voice authentication are the following: user's voice variations (voice can vary depending on health conditions, age, mood etc.); presence of a noise component. Solving these problems will allow to use the voice authentication technology to ensure the best protection of personal data, ease of usage. Besides, it is the cheapest among the existing identification technologies.

Materials and methods. The authors used numerical methods, digital signal processing, spectral methods, methods of mathematical statistics and time-series, as well as artificial intelligence and pattern recognition. A fragmentating device is based on a continuous-discrete model of speech processing, which, in combination with a narrow-band filter, allows to determine the average duration of sound.

Results. It is shown that the qualitative tone/noise speech qualifier should give the output of "0" and "1", the duration of which is described by the continuum-discrete distribution of values of duration of the intervals between sections of tonal sounds, distributed according to normal laws. The discrete part of the distribution is formed by the discrete nature of the flow of tone sounds and noise appearance in speech, as well as their combinations (pairs, triples, quadruples, and etc). The continuous (continuum) part of sound length values distribution is conditioned by the instability of speech by changing a pronunciation pace. The article describes a method of calculating the average length of one sound of a meaningful speech. This study has allowed to build a machine to determine the average length of sound in different parts of an audio signal.

Conclusions. The article suggests a numerical algorithm for identification of individual speaker's speech, allowing to sync speech areas. Usage of the developed algorithm has allowed to specify parameter values that characterize the statistical description of duration of the intervals between speech sounds and noise between tonal sounds. The study has made it possible to build a machine to determine the average length of sound in different parts of an audio signal. The results are the basis for building neural network authentication technologies.

Key words: numerical methods, digital signal processing, biometric systems, voice authentication.

Введение

В настоящее время активно идут процессы информатизации современного общества. В России и за рубежом объявлено о создании электронного правительства. Каждая из федеральных служб России (Налоговая инспекция, ГИБДД, Пенсионный фонд РФ и др.) создают на своих сайтах личные кабинеты для взаимодействия с гражданами РФ. Каждый личный кабинет должен быть защищен уникальным логином и длинным паролем из случайных знаков. К сожалению, многие пользователи не могут запомнить множество логинов и множество длинных паролей доступа из 32 случайных знаков.

По мнению экспертов, выход из создавшегося положения состоит в использовании преобразователей личной биометрии человека в его логин и длинный пароль доступа. В США и странах Евросоюза эту проблему решают с помощью «нечетких экстракторов» [1, 2]. В России и Казахстане развивается технология использования больших искусственных нейронных сетей [3, 4].

Для преобразования в логин и код пароля доступа может быть использована любая биометрическая технология, однако для доступа в личные кабинеты желательно использовать анализ особенностей голоса, так как микрофон является самым распространенным и самым простым средством ввода биометрии.

Обычно для преобразователя голосовой биометрии в контролируемые параметры используют линейные предсказатели [5, 6]. Линейный предсказатель строится на основе анализа нескольких рядом расположенных отсчетов изменения значений давления волны звука, по которым предсказывается следующее значения отсчета.

Линейные предсказатели работают по следующему алгоритму. Пусть на линейную модель системы воздействует сигнал возбуждения $x(n)$, а на ее выходе формируется сигнал $y(n)$. Параметры системы неизвестны, а вход и выход связаны разностным уравнением

$$y(n) = b_0x(n) - \sum_{k=1}^{M-1} a_k y(n-k), \quad n = M, M+1, \dots, \quad (1)$$

где b_0 – известный коэффициент, а коэффициенты a_k , $k = 1, 2, \dots, M-1$, неизвестны.

Требуется определить коэффициенты $\{a_k\}$, $k = 1, 2, \dots, M-1$.

Нахождение коэффициентов системы уравнений (1) сводится к решению линейной системы уравнений

$$y(l) = b_0x(l) - \sum_{k=1}^{M-1} a_k y(l-k), \quad l = 1, 2, \dots, M-1,$$

если известны начальные значения

$$y(0), y(-1), \dots, y(-M+2).$$

Если начальные значения неизвестны, то задача сводится к решению линейной системы уравнений

$$y(n) = b_0x(n) - \sum_{k=1}^{M-1} a_k y(n-k), \quad n = M, M+1, \dots, 2M-1,$$

относительно неизвестных значений $\{a_k\}$.

В случае предсказания голосовых сигналов используется следующий алгоритм. Выделяется фраза (или ее фрагмент) и разбивается на кадры. Каждые два следующие друг за другом кадры связаны системой разностных уравнений

$$s(n) = \sum_{k=1}^L a_k s(n-k) + \varepsilon_n, \quad n = L+1, L+2, \dots, 2L, \quad (2)$$

где ε_n – погрешность; s_n – значения сигнала $s(t)$ на выбранной сетке узлов.

Для решения системы (2) относительно коэффициентов $\{a_k\}$, $k = 1, 2, \dots, L$, можно использовать численные методы.

В результате решения системы уравнений (2) на последовательности кадров каждому кадру начиная с первого ставится в соответствие вектор $A^l = \{a_1^l, a_2^l, \dots, a_L^l\}$, где l – номер кадра. Из векторов A^l , $l=1, 2, \dots$, составляются различные логины, идентифицирующие диктора.

Алгоритмы идентификации звуковых сигналов

Линейный предсказатель звуковых сигналов определяется как линейная система, на выходе которой имеем

$$s(n) = \sum_{k=1}^{12} a_k s(n-k) + e(n), \quad (3)$$

где $a_1, a_2, a_3, \dots, a_l$ – неизвестные коэффициенты линейного предсказания в пределах обрабатываемого сегмента речевого сигнала; $e(n)$ – ошибка предсказания; $s(n)$ – значение отсчета звука, предсказанное по l предыдущим отсчетам; n – текущий номер отсчета обрабатываемого звукового файла.

1. В качестве значения обрабатываемого сегмента выбирается число, при котором удастся решать систему линейных уравнений в реальном времени. Так, в стандартном вокодере LPC-12 используется 12 коэффициентов линейного предсказания. Число коэффициентов линейного предсказателя не может быть велико, так как обусловленность решаемой системы линейных уравнений ухудшается с ростом ее размерности.

Ниже рассматривается линейный предсказатель с коэффициентами $\{\alpha_k\}$, который определяется как линейная система с заданным параметром $l=12$.

Для преодоления негативных свойств вычислительных процессов, связанных с плохой обусловленностью матриц линейных предсказателей, ниже предлагаются и обосновываются итерационные методы.

2. В линейных системах параметры $\{\alpha_k\}$ линейного предсказателя находят, минимизируя ошибку $e(n)$:

$$\sum_{m=1}^{12} s(n+m-i)s(n+m) = \sum_{k=1}^{12} a_k \sum_{m=1}^{12} s(n+m-i)s(n+m-k), \quad i \leq i \leq 12. \quad (4)$$

Система уравнений (3) является линейной относительно неизвестных коэффициентов $\{\alpha_k\}$, $k=1, 2, \dots, 12$, системой алгебраических уравнений.

Введем обозначения:

$$\sum_{m=1}^{12} s_n(m-i)s_n(m-k) = R_{ik}^n,$$

$$\sum_{l=1}^{12} s_n(m-i)s_n(m) = R_i^n,$$

где R_{ik}^n – коэффициенты автокорреляции.

3. Запишем систему (4) в виде

$$\sum_{k=1}^{12} \alpha_k R_{ik}^n = R_i^n, \quad i = 1, 2, \dots, n. \quad (5)$$

Замечание. Выше было использовано 12 отчетов, так как эта размерность принята в большинстве технических реализаций линейных предсказателей. Отметим, что результаты, полученные в данной работе, справедливы для произвольного L числа отсчетов.

Система уравнений (5) является системой с теплоцевой матрицей. Как известно, теплоцевы матрицы плохо обусловлены и решение системы уравнений (5) с экспериментально полученными данными методами линейной алгебры может привести к значительным ошибкам.

Более предпочтительными являются итерационные методы.

Так как $R_{ik} = R_{ki}$, $1 \leq k, i \leq 12$, то матрица системы (5) является самосопряженной. Напомним, что матрица M называется самосопряженной, если $MM^* = M^*M$, где M^* – матрица, сопряженная матрице M .

Изложим итерационный метод решения систем линейных алгебраических уравнений с самосопряженными матрицами. Напомним, что оператор K называется самосопряженным, если $KK^* = K^*K$, где K^* – оператор, сопряженный оператору K .

Для общности рассуждений рассмотрим операторное уравнение

$$Kx = f \quad (6)$$

с самосопряженным оператором K в гильбертовом пространстве H ($K \in [H, H]$).

Известно [7], что спектр самосопряженного оператора K неотрицательный ($0 \leq \sigma(K)$). Здесь через $\sigma(K)$ обозначен спектр оператора K . Пусть $\gamma = 1/(2\|K\|)$, где $\|K\|$ – норма оператора K , и уравнению (6) поставим в соответствие уравнение

$$\gamma Kx = \gamma f. \quad (7)$$

Уравнение (7) будем решать итерационным методом:

$$x_{n+1} = \lambda_n x_n + (1 - \lambda_n)(x_n - \gamma(Kx_n - f)), \quad (8)$$

$n = 0, 1, \dots$; $0 < \lambda_* \leq \lambda_n \leq \lambda^* < 1$. Естественно положить $\lambda_n = 1/2, n = 0, 1, \dots$

Справедливо следующее утверждение.

Утверждение 1. Пусть K – самосопряженный оператор в рефлексивном пространстве H . Тогда итерационный процесс (8) сходится к решению уравнения (7) и, следовательно, к решению уравнения (6). Если известно, что спектр оператора K положительный ($0 < \beta < \sigma(K)$), то итерационный процесс (8) сходится к решению x^* уравнения (5) со скоростью геометрической прогрессии со знаменателем $q = (1 - (1 - \lambda_*)\beta/2)$ и справедлива оценка $\|x^* - x_{n+1}\| = q^n \|x_1 - x_0\|$.

В случае, если $0 \in \sigma(K)$, то задача решения системы уравнений (7) является некорректной и необходимо использовать методы регуляризации. Остановимся на регуляризации по Тихонову. В этом случае естественно реализовать следующий итерационный процесс:

$$x_{n+1} = \lambda_n x_n + (1 - \lambda_n) (x_n - \gamma(\beta x_n + Kx_n - f)), \quad n = 0, 1, \dots, n_x, n, \dots, \quad (9)$$

где $0 < \lambda_* \leq \lambda_n \leq \lambda^* < 1$, $\beta > 0$ – параметр регуляризации.

Сходимость итерационного процесса (9) доказывается по аналогии с доказательством сходимости итерационного процесса (8).

Возвращаясь к системе уравнений (2), отметим, что к ней применимы итерационные схемы (8) и (9).

Представляет интерес сравнение итерационной схемы (8) для системы уравнений (2) с известным методом Левинсона – Дарбина [8].

Нетрудно видеть, что число арифметических действий при реализации итерационной схемы (8) для решения уравнения $Kx = f$, где $\min \sigma(K) > 0$, значительно меньше, чем в методе Левинсона – Дарбина. Кроме того, метод (8) дает гарантированную оценку погрешности. Из проведенных рассуждений следует перспективность использования итерационных методов в случае матриц большой размерности.

Сходимость вычислительных схем (8)–(9) доказана в рефлексивных пространствах. Среди наиболее распространенных пространств к рефлексивным пространствам относятся гильбертовы пространства L_2 , l_2 и евклидово пространство E_n . В частности, в случае линейного предсказателя уравнение (4) решается в пространстве E_{12} векторов $A = (\alpha_1, \dots, \alpha_{12})$ с нормой

$$\|A\| = \left[\sum_{k=1}^{12} |\alpha_k|^2 \right]^{1/2}.$$

Отметим, что итерационная схема (9) используется в случае, когда нуль принадлежит спектру оператора K . В этом случае вводится дополнительный регуляризирующий оператор βx , который «сдвигает» спектр оператора K вправо. Выбор параметра регуляризации представляет собой сложную задачу, требующую анализа априорной информации об исследуемом процессе.

Поэтому представляет интерес рассмотрение других методов решения уравнений на спектре. В частности, для решения уравнения на спектре может быть использован метод Бакушинского – Страхова [9].

Рассмотрим операторное уравнение

$$K_1 x = f_1, \quad (10)$$

где $K_1 = \gamma K$, $f_1 = \gamma f$, $\gamma = 1 / \|K\|$.

В этом случае $\|K_1\| = 1$ и, так как оператор K_1 – самосопряженный, то его спектр лежит на сегменте $[0, 1]$. Из теоремы Данфорда следует, что спектр оператора $E - K_1$, где E – тождественный оператор, лежит на сегменте $[0, 1]$. Так как оператор $E - K_1$ самосопряженный, то $\|E - K_1\| = 1$.

Следовательно, для оператора K_1 выполнены условия $\|K_1\| = 1$, $\|E - K_1\| = 1$.

При этих условиях в рефлексивном банаховом пространстве справедливо следующее утверждение.

Утверждение 2 [9]. Пусть $x_0 \in B$ – произвольный элемент,

$$x_{n+1} = (E - K_1)x_n + f_1,$$

$$\bar{x}_n = \frac{1}{n+1} \sum_{k=0}^n x_k. \quad (11)$$

Последовательность \bar{x}_n сходится к решению x^* уравнения (11).

Следствие. Последовательность (11) сходится к решению уравнения (10).

4. В данной работе параметры $\{\alpha_k\}$ получают, минимизируя ошибку $e(n)$. В результате приходим к системе уравнений

$$\sum_{m=1}^{12} s_n(m-i)s_n(m) = \sum_{k=1}^{12} a_k \sum_{m=1}^{12} s_n(m-i)s_n(m-k), i \leq i \leq 12. \quad (12)$$

Алгоритм, осуществляющий классификацию участков звука на тональные и шумовые участки речи одного человека без шумов, использует вычисление автокорреляционной функции ошибки линейного предсказателя $e(n)$. На рис. 1 приведен пример автокорреляционной функции ошибок линейного предсказания.

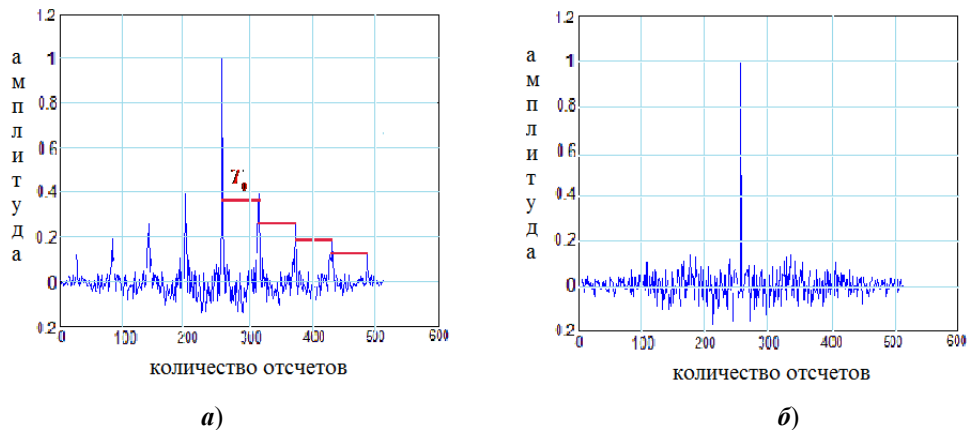


Рис. 1. Автокорреляционная функция на тональном участке звука (а), шумовом участке голосовой фразы (б)

Из рис. 1 видно, что автокорреляционные функции ошибки линейного предсказания для тональных участков звука и для шумовых участков звука существенно отличаются. В автокорреляционной функции тональных участков имеются повторяющиеся пики, расстояние T_0 между которыми соответ-

стует периоду основного тона обрабатываемой речи человека. Шумовые участки речи человека не содержат периодической составляющей (смотри правую часть рис. 1).

Таким образом, классификация звуков речи одного человека (без посторонних шумов) на тональные и шумовые участки является уже решенной задачей. Для выделения тональных участков достаточно обнаружить периодичность пиков автокорреляционной функции ошибки линейного предсказателя. Однако как только появляется смесь речи двух или более людей или появляется значительная шумовая составляющая, линейный предсказатель перестает работать.

Дискретное статистическое описание длительности интервалов между шумовыми звуками речи и между тональными звуками

Предположим, что создан идеальный читающий робот, который читает осмысленный текст на русском языке. Будем считать, что он каждую букву текста будет воспроизводить шумовыми и тональными звуками одинаковой длительности. Кроме того, будем считать, что пробел между словами и предложениями текста будет воспроизводиться говорящим роботом как пауза, заполняемая тихим внешним шумом. В этом случае для русского языка интервалы между последовательностями тональных звуков будут описываться дискретным распределением, пример которого приведен на рис. 2.

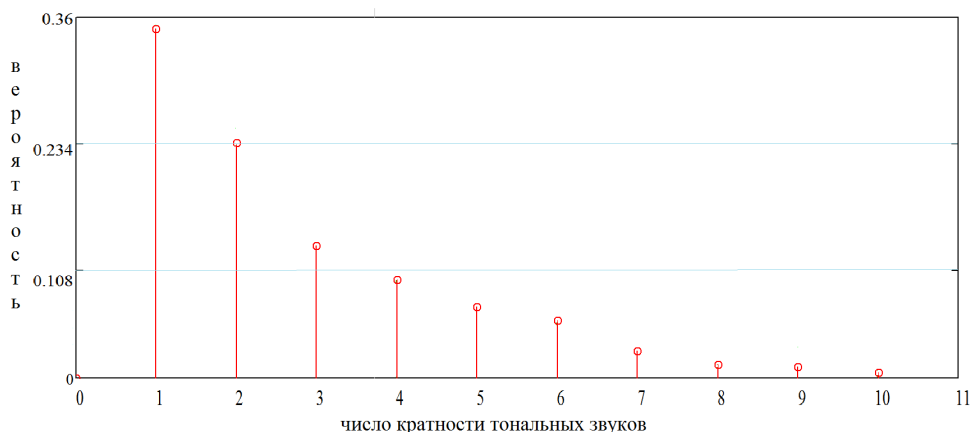


Рис. 2. Дискретное распределение значений длительности тональных звуков русской речи для монотонно читающего текст робота

Приведенное на рис. 2 дискретное распределение получено исследованием случайного русского текста, состоящего из 15 500 знаков. Шипящими звуками являются следующие фонемы: «ф», «х», «ц», «б», «п», «к», «т», «ь», «ъ», «с», «ч», «щ», «ш». При анализе текста учтены паузы. По оси y отложена вероятность появления кратных тональных звуков (число идущих подряд тональных звуков). По оси x отложена кратность тональных звуков. Как видно из рис. 2, наиболее часто встречаются одиночные тональные звуки, т.е. одна тональная фонема, находящиеся между различными вариантами шипящих. Можно сделать вывод, что вероятность появления двух тональных звуков подряд меньше чем одинарных, вероятность трех тональных меньше чем ве-

роятность парных и т.д. Максимальное число подряд стоящих тональных звуков из тестового текста не превышает десяти.

Получение классификации тон/шум на практике является исходной задачей при построении более экономичных и качественных вокодеров, а также в создании голосовой идентификации (аутентификации) пользователя. Одним из направлений совершенствования классификатора тон/шум является использование узкополосных, перестраиваемых фильтров частоты основного тона, подстраиваемого под период основного тона тестируемого диктора. В данном исследовании тестирование проходило при высокой частоте дискретизации, равной 44100 Гц. Период основного тона при данной частоте может варьироваться, поэтому требуется постоянное его уточнение на каждом участке звука. Выделение тональных участков происходило за счет обращения шума в ноль на тех участках, на которых выходной сигнал не приобретал синусоидальную форму, либо (в случае полного несоответствия) приобретал нулевое значение. Преимущество применения данного фильтра в классификации тон/шум заключается в устойчивой работе по выделению периода основного тона и возможности отделения шумовых звуков и шумовых компонент сигнала. На рис. 3 приведен пример распределения длительности тональных участков речи, полученного на базе из 500 образов. Данная гистограмма показывает, что дисперсия каждого из распределений (общей длины подряд идущих тональных звуков), полученная сочетанием различного числа звуков, примерно кратна дисперсии длины одинарного звука. Отсюда вытекает возможность в выделении средней длины звука речи отдельного человека.

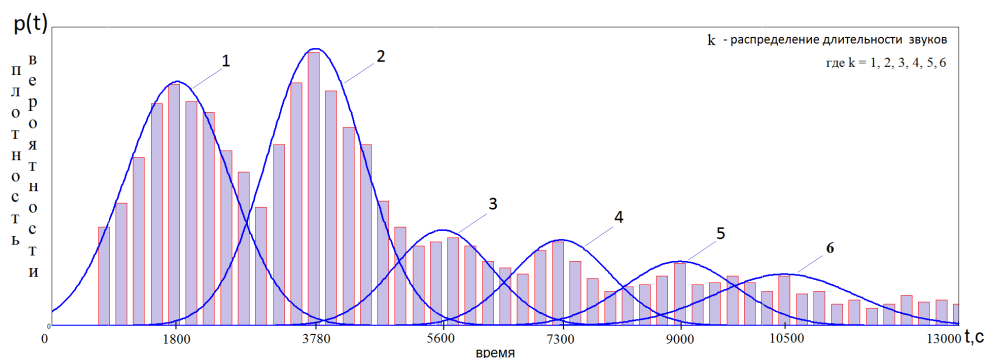


Рис. 3. Дискретно-континуальное распределение длительности интервалов между участками тональных звуков, распределенных по нормальным законам

Для той же базы было построено распределение длин шумовых звуков. Как показывает гистограмма (рис. 4), длительность кратных звуков (идущих подряд звуков) примерно кратна длительности одиночных звуков. Например, в распределении длительности двоичных звуков примерно содержится длина двух одинарных и т.д.

Рисунки 3 и 4 приведены к равному масштабу для сравнения длительности тональных и шумовых участков речи. При наложении распределения друг на друга можно заметить кратное соответствие в дисперсии длительности тональных и шумовых звуков (13). Данное утверждение говорит о правильной работе классификатора тон/шум и о возможности выделения необходимого биометрического параметра – средней длины звука:

$$\sigma_{\text{шум}} = \sigma_{\text{тон}} \quad (13)$$

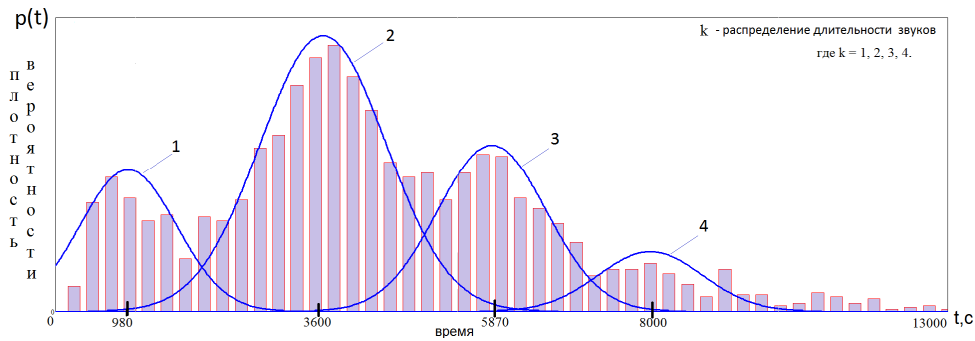


Рис. 4. Дискретно-континуальное распределение значений длительности шумовых звуков русской речи

Как уже было замечено, математическое ожидание всплесков кратно друг другу как для тональных, так и для шумовых звуков. Соответственно в первую очередь должна решаться задача нахождения данной кратности:

$$\begin{cases} E(T_1) = C_{\text{Био}}; \\ E(T_2) = 2C_{\text{Био}}; \\ \dots\dots\dots \\ E(T_n) = nC_{\text{Био}}, \end{cases} \quad (14)$$

где $C_{\text{Био}}$ – средняя длительность звука.

Средняя длительность звука является величиной статистической и может варьироваться при смене темпа речи, поэтому после нахождения кратности вычисляется данный параметр. Для определения длительности возможных вариантов сочетаний звуков должен осуществляться непрерывный подсчет данного параметра согласно системе

$$\begin{cases} E_{\text{тон}}(T_n) = (n-1)E_{\text{тон}}(T_{n-1}) = (n-2)E_{\text{тон}}(T_{n-2}) = \dots = nE_{\text{тон}}(T_1); \\ E_{\text{шум}}(T_n) = (n-1)E_{\text{шум}}(T_{n-1}) = (n-2)E_{\text{шум}}(T_{n-2}) = \dots = nE_{\text{шум}}(T_1); \\ \sigma = \frac{(E_{\text{шум}}(T_1) + E_{\text{тон}}(T_1))}{2}. \end{cases} \quad (15)$$

На рис. 5 представлено исследование пауз на той же тестируемой базе. Распределение не показало никакой закономерности в гистограмме при корреляции тональных и шумовых звуков.

Данный параметр необходим в получении наиболее устойчивой синхронизации при выделении индивидуальных голосовых параметров человека.

В первую очередь параметр средней длины звука необходим для синхронизации классификатора тон/шум при повторном произношении парольной голосовой фразы. На рис. 6 представлен пример работы синхронизации при 9-кратном произношении одной и той же фразы, принадлежащей одному диктору.

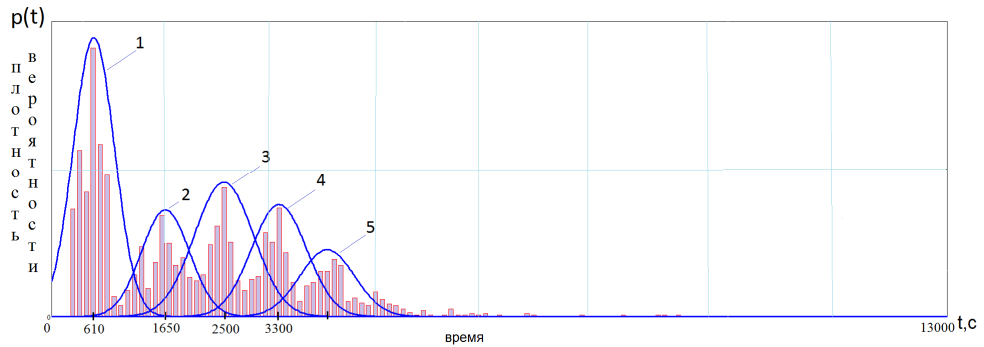


Рис. 5. Дискретное распределение длительности интервалов между участками пауз и шумовых/тональных звуков для русской речи

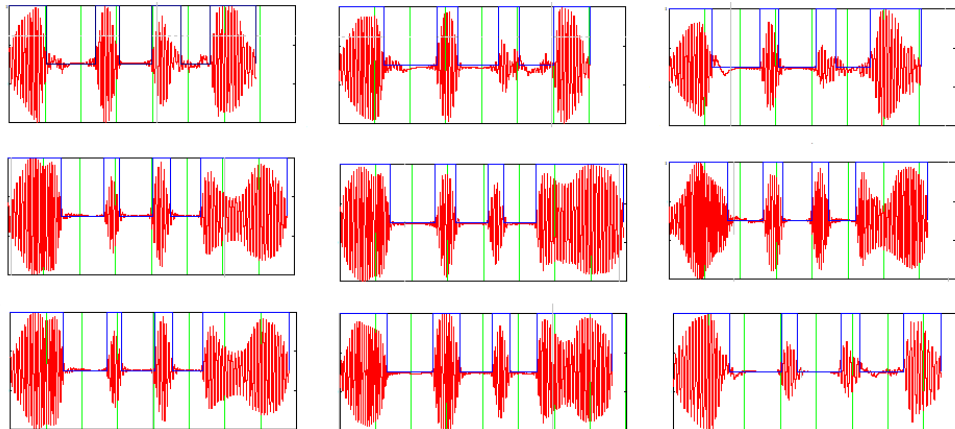


Рис. 6. Графики выходных сигналов после процесса синхронизации классификатора тон/шум на голосовой фразе

На графиках изображен сигнал, прошедший через узкополосный фильтр и построены прямые, огибающие данный сигнал. Прямые получены в результате классификации тон/шум (участки, на которых выявлена детерминированная часть сигнала, представлены единицей, участки, на которых выявлен шум, представлены нулем).

Как видно из рис. 6, длина каждого тонального участка варьируется в различных пределах. Значение каждого из участков уточняется при добавлении речевых образцов. В каждом звуковом фрагменте может содержаться несколько фонем. Точное определение количества взятых образцов зависит от точности определения средней длины звука.

На рис. 7 представлен пример конечного процесса синхронизации фонем в обрабатываемой голосовой фразе.

Приведем доказательство утверждения 1.

Доказательство сходимости итерационного процесса (8) основано на следующем утверждении [10]. Пусть B – рефлексивное банахово пространство, K – линейный оператор из B в B и $\|K\| = 1$.

Замечание. Евклидово пространство векторов является рефлексивным.

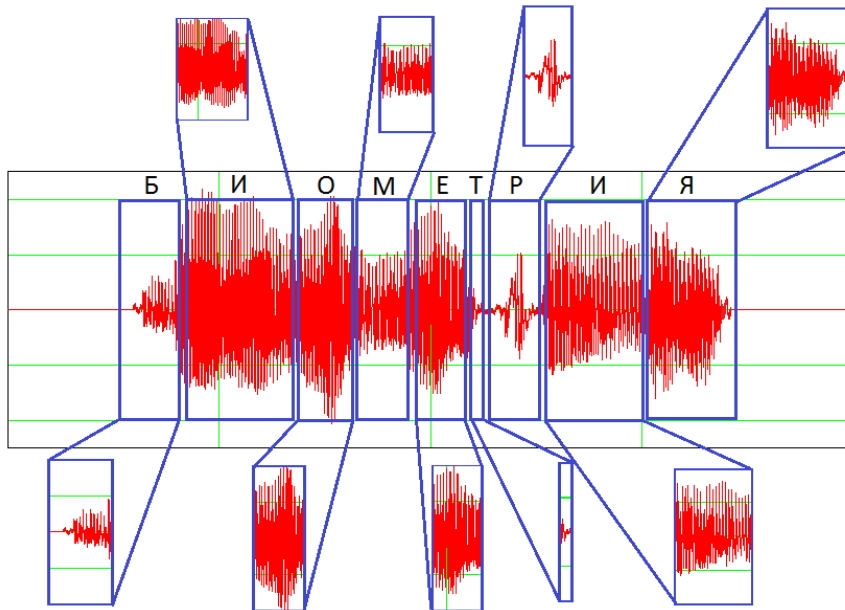


Рис. 7. Фрагментация заданного парольного слова

Итерационный процесс

$$x_{n+1} = \lambda_n x_n + (1 - \lambda_n) Kx_n, \quad 0 < \lambda_* \leq \lambda_n \leq \lambda^* < 1, \quad n = 0, 1, \dots,$$

сходится к решению x^* уравнения $Kx = 0$, если последнее разрешимо.

Для применимости теоремы 1 к доказательству сходимости итерационного процесса (8) нужно показать, что $\|I - \gamma K\| \leq 1$. Так как оператор K самосопряженный, то его спектр находится на положительной полуоси (включая, возможно, начало координат). Так как $\gamma = 1 / (2\|K\|)$, то спектр γK расположен в сегменте $[0, 1/2]$, а по теореме Данфорда [11] спектр оператора $I - \gamma K$ расположен в сегменте $[\frac{1}{2}, 1]$. Следовательно, $\|I - \gamma K\| = 1$ и условия теоремы 1 выполнены. Итерационный процесс (8) сходится.

В случае, если известно, что спектр оператора K положительный ($0 < \beta < \sigma(K)$), то можно получить более сильное утверждение.

Из проведенных выше утверждений следует, что спектр оператора $I - \gamma K$ расположен в сегменте $[\frac{1}{2}, 1 - \frac{\beta}{2}]$. Следовательно, $\|I - \gamma K\| = 1 - \frac{\beta}{2}$. Это позволяет оценить скорость сходимости итерационного процесса (8). В самом деле,

$$\begin{aligned} \|x_{n+1} - x_n\| &\leq \lambda_n \|x_n - x_{n-1}\| + (1 - \lambda_n) \|(x_n - x_{n-1} - \gamma K(x_n - x_{n-1}))\| \leq \\ &\leq \lambda_n \|x_n - x_{n-1}\| + (1 - \lambda_n) \left(1 - \frac{\beta}{2}\right) \|x_n - x_{n-1}\| \leq \end{aligned}$$

$$\leq 1 - (1 - \lambda_n) \frac{\beta}{2} \|x_n - x_{n-1}\|, n = 0, 1, \dots \quad (16)$$

Из теоремы Банаха [12] и неравенства (9) следует, что итерационный процесс (8) сходится к решению x^* уравнения (6) со скоростью геометрической прогрессии со знаменателем $q = (1 - (1 - \lambda_*)\beta / 2)$ и справедлива оценка

$$\|x^* - x_{n+1}\| = q^n \|x_1 - x_0\|.$$

Заключение

При применении построенного алгоритма кластеризации звуковой фразы стало возможным синхронизировать и кластеризовать участки речи с наибольшей точностью. Учитывая параметры, описанные в статье, удалось построить нейросетевую технологию распознавания речи, основанную на анализе как в спектральной, так и временной области параметров строгой последовательности фрагментированных участков речи.

Для каждого фрагментированного участка речи строится нейронная сеть, которая вырабатывает биологический ключ для каждой фонемы. Вектор, составленный из биологических ключей, является идентификатором конкретной личности.

Полученное исследование позволило построить автомат по выделению средней длины звука на различных участках звукового сигнала.

Список литературы

1. **Dodis, Y.** Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy / Y. Dodis, L. Reyzin, A. Smith // EUROCRYPT. – 2004, April 13. – P. 523–540.
2. **Monrose, F.** Cryptographic key generation from voice / F. Monrose, M. Reiter, Q. Li, S. Wetzel // Proc. IEEE Symp. on Security and Privacy, 2001.
3. Нейросетевая защита персональных биометрических данных / Ю. К. Язов, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, И. Г. Назаров. – М. : Радиотехника, 2012. – 157 с.
4. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа : моногр. / Б. С. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы : Изд-во LEM, 2014. – 144 с. – URL: <http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf>
5. **Рамишвили, Г. С.** Автоматическое опознавание говорящего по голосу / Г. С. Рамишвили. – М. : Радио и связь, 1981. – 224 с.
6. **Маркел, Дж. Д.** Линейное предсказание речи / Дж. Д. Маркел, А. Х. Грей. – М. : Радио и связь, 1980. – 248 с.
7. **Канторович, Л. В.** Функциональный анализ / Л. В. Канторович, Г. П. Акилов. – М. : Наука, 1977. – 750 с.
8. **Соломина, А. И.** Основы цифровой обработки сигналов / А. И. Соломина, Д. А. Улахович, С. М. Арбузов, Е. Б. Соловьева. – СПб., 2013. – 768 с.
9. **Бакушинский, А. Б.** О решении некоторых интегральных уравнений 1 рода методом последовательных приближений / А. Б. Бакушинский, В. Н. Страхов // Журнал вычислительной математики и математической физики. – 1968. – Т. 8, № 1. – С. 181–185.

10. **Обломская, Л. Я.** О методах последовательных приближений для линейных уравнений в банаховых пространствах / Л. Я. Обломская // Журнал вычислительной математики и математической физики. – 1968. – Т. 8, № 2. – С. 417–426.
11. **Данфорд, Н.** Линейные операторы. Т. 1. Общая теория / Н. Данфорд, Дж. Шварц. – М.: ИЛ, 1962. – 895 с.
12. **Люстерник, Л. А.** Элементы функционального анализа / Л. А. Люстерник, В. И. Соболев. – М.: Наука, 1965. – 540 с.

References

1. Dodis Y., Reyzin L., Smith A. *EUROCRYPT*. 2004, April 13, pp. 523–540.
2. Monrose, F., Reiter M., Li Q., Wetzels S. *Proc. IEEE Symp. on Security and Privacy, 2001*.
3. Yazov Yu. K., Volchikhin V. I., Ivanov A. I., Funtikov V. A., Nazarov I. G. *Neurosetevaya zashchita personal'nykh biometricheskikh dannykh* [Neural network protection of personal biometric data]. Moscow: Radiotekhnika, 2012, 157 p.
4. Akhmetov B. S., Ivanov A. I., Funtikov V. A., Bezyaev A. V., Malygina E. A. *Tekhnologiya ispol'zovaniya bol'shikh neyronnykh setey dlya preobrazovaniya nechetkikh biometricheskikh dannykh v kod klyucha dostupa: monogr.* [Technology of using large neural networks for fuzzy biometric data conversion to access key codes: monograph]. Almaty: Izd-vo LEM, 2014, 144 p. Available at: <http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf>
5. Ramishvili G. S. *Avtomaticheskoe opoznavanie govoryashchego po golosu* [Automatic voice authentication of speakers]. Moscow: Radio i svyaz', 1981, 224 p.
6. Markel Dzh. D., Grey A. Kh. *Lineynoe predskazanie rechi* [Linear voice prediction]. Moscow: Radio i svyaz', 1980, 248 p.
7. Kantorovich L. V., G. P. Akilov *Funktsional'nyy analiz* [Functional analysis]. Moscow: Nauka, 1977, 750 p.
8. Solomina A. I., Ulakhovich D. A., Arbuzov S. M., Solov'eva E. B. *Osnovy tsifrovoy obrabotki signalov* [Basic digital signal processing]. Saint-Petersburg, 2013, 768 p.
9. Bakushinskiy A. B., Strakhov V. N. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki* [Journal of calculus mathematics and mathematical physics]. 1968, vol. 8, no. 1, pp. 181–185.
10. Oblomskaya L. Ya. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki* [Journal of calculus mathematics and mathematical physics]. 1968, vol. 8, no. 2, pp. 417–426.
11. Danford N., Shvarts Dzh. *Lineynye operatory. T. 1 Obshchaya teoriya* [Linear operators]. Moscow: IL, 1962, 895 p.
12. Lyusternik L. A., Sobolev V. I. *Elementy funktsional'nogo analiza* [Functional analysis elements]. Moscow: Nauka, 1965, 540 p.

Бойков Илья Владимирович

доктор физико-математических наук,
профессор, заведующий кафедрой
высшей и прикладной математики,
Пензенский государственный
университет (Россия, г. Пенза,
ул. Красная, 40)

E-mail: boikov@pnzgu.ru

Boikov Ilya Vladimirovich

Doctor of physical and mathematical
sciences, professor, head of sub-department
of higher and applied mathematics, Penza
State University (40 Krasnaya street,
Penza, Russia)

Иванов Александр Иванович

доктор технических наук, начальник
лаборатории биометрических
и нейросетевых технологий,
Пензенский научно-исследовательский
электротехнический институт (Россия,
г. Пенза, ул. Советская, 9)

E-mail: ivan@pniei.ru

Ivanov Aleksandr Ivanovich

Doctor of engineering sciences, head
of laboratory of biometric and neural
network technologies, Penza Research
Institute of Electrical Engineering
(9 Sovetskaya street, Penza, Russia)

Калашников Дмитрий Михайлович

аспирант, Пензенский
государственный университет
(Россия, г. Пенза, ул. Красная, 40)

E-mail: kalashnikovdm.penza@gmail.com

Kalashnikov Dmitriy Mikhaylovich

Postgraduate student, Penza State
University (40 Krasnaya street,
Penza, Russia)

УДК 004; 519.7; 519.6; 519.66; 612.087.1

Бойков, И. В.

Алгоритм построения статистического дискретно-континуального описания длительности звуков потока осмысленной речи диктора / И. В. Бойков, А. И. Иванов, Д. М. Калашников // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2015. – № 4 (36). – С. 64–78.