

**ИССЛЕДОВАНИЕ ПРИЗНАКОВОГО ПРОСТРАНСТВА
ДЛЯ ОПИСАНИЯ ОТРЕЗКОВ РЕЧЕВЫХ СИГНАЛОВ
В ЗАДАЧАХ РАСПОЗНАВАНИЯ УСТНОЙ РЕЧИ**

© 2020 Е. Г. Жилияков¹, А. В. Болдышев^{✉2}, С. В. Уманец²

¹*Белгородский государственный национальный исследовательский университет,
ул. Победы, 85, 308015 Белгород, Российская Федерация*

²*Белгородский филиал ПАО «Ростелеком»,
пр-кт Б.-Хмельницкого, 81, 308009 Белгород, Российская Федерация*

Аннотация. В данной работе рассматривается набор признаков для описания отрезков речевых сигналов с точки зрения их информативности. Наборы признаков используются в задачах распознавания. От информативности набора признаков во многом зависит успех решения задач распознавания. Информативность набора признаков оценивается с помощью информационной энтропии. В данной работе производится сравнение информационной энтропии для наборов признаков, полученных с помощью метода Фурье и субполосного метода. Также рассматриваются модификации этих методов для анализа речевых сигналов, а именно переход к мел шкале. Кроме мел шкалы, для субполосного метода рассматриваются и другие искажения частотной оси. В работе приводятся результаты вычисления информационной энтропии для разных способов получения набора признаков, описывающих отрезки речевых сигналов. В результате делается вывод о повышении информативности набора признаков при использовании субполосного метода с нелинейным разбиением частотной оси на полосы анализа.

Ключевые слова: речевой сигнал, признаки, субполосный метод, мел шкала, частотная ось, доли энергии, энтропия.

ВВЕДЕНИЕ

В задачах, связанных с обработкой речевых сигналов (распознавание [1], синтез речи, сжатие [2]) необходимо сформировать определенные признаки. Эти признаки должны быть информативными — т. е. позволять эффективно решать поставленную задачу. При этом они не должны обладать избыточностью, так как это может повлиять на точность алгоритмов или скорость их выполнения. В случае с речевыми сигналами наиболее часто проводится анализ именно частотных

свойств рассматриваемого сигнала. Речевой сигнал — это результат регистрации электромагнитных колебаний на выходе микрофона от воздействия акустических колебаний во время разговора. В результате анализа получается спектр — ряд значений, поставленный в соответствие частотной шкале. Значения полученного спектра и представляют собой вектор признаков, которые используются для решения задач распознавания, фильтрации, сжатия и ряда других. Оценить избыточность признаков можно с помощью информационной энтропии [3].

Цель данной работы является исследование признакового пространства для описания отрезков речевых сигналов в задачах рас-

✉ Болдышев Алексей Владимирович
e-mail: alexeiboldyshev@mail.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

познавания устной речи на основе сравнения информационной энтропии. Предполагается, что сравнению подвергаются признаковые пространства, полученные методом Фурье и субполосным методом [4].

1. МЕТОДЫ ИССЛЕДОВАНИЯ

Информационная энтропия для набора из $r = 1, 2, \dots, R$ признаков определяется выражением [3]:

$$H = -\sum_{r=1}^R P_r \log(P_r), \quad (1)$$

где P_r — вероятность появления r -го признака.

Максимальное значение энтропии у равномерно распределённых признаков, когда $P_r = 1/R$. Величина энтропии при этом $H = \log(R)$. Чтобы исключить зависимость от количества признаков R при сравнении разных наборов признаков, предлагается ввести нормирующий множитель $\frac{1}{\log(R)}$, т. е. вычислять значение информационной энтропии так:

$$H = -\frac{1}{\log(R)} \sum_{r=1}^R P_r \log(P_r). \quad (2)$$

Для анализа оцифрованных сигналов часто применяется частотный анализ в базисе Фурье. Для этого используется [5] дискретным прямым (2) и обратным (3) преобразованиями Фурье. Для анализируемого отрезка сигнала, длительностью N отсчетов x_i , где $i = 1, 2, \dots, N$, коэффициенты ряда Фурье определяются следующим образом:

$$X_k = \sum_{i=1}^N x_i e^{-j \frac{2\pi}{N} (i-1)(k-1)}, \quad (3)$$

где $k = 1, 2, \dots, N$, j — комплексная единица, $j = \sqrt{-1}$.

Обратное преобразование:

$$x_i = \frac{1}{N} \sum_{k=1}^N X_k e^{j \frac{2\pi}{N} (i-1)(k-1)}. \quad (4)$$

Вещественная часть трансформанты Фурье (косинусная) определяется так:

$$\text{Re}(X_k) = X_k \cos\left(\frac{2\pi}{N} k\right), \quad (5)$$

мнимая (синусная) так:

$$\text{Im}(X_k) = X_k \sin\left(\frac{2\pi}{N} k\right). \quad (6)$$

Затем определяются по формулам следующие значения:

амплитуда

$$A_k = \sqrt{\text{Re}(X_k)^2 + \text{Im}(X_k)^2} \quad (7)$$

фаза

$$\Psi_k = \text{arctg}\left(\frac{\text{Im}(X_k)}{\text{Re}(X_k)}\right) \quad (8)$$

энергия

$$W_k = \text{Re}(X_k)^2 + \text{Im}(X_k)^2. \quad (9)$$

В качестве информационных признаков в данной работе будут рассматриваться нормированные значения энергии — доли энергии:

$$P_k = \frac{W_k}{\sum_{k=1}^{\lfloor N/2 \rfloor} W_k}, \quad (10)$$

где $k = 1, 2, \dots, \lfloor \frac{N}{2} \rfloor$ из-за симметричности ряда Фурье относительно своей середины, квадратные скобки — это взятие целой части.

Другой способ вычисления распределения энергии по частотам — субполосный метод (здесь T — знак транспонирования)

$$W_r = \vec{x} \cdot \mathbf{A}_r \cdot \vec{x}^T. \quad (11)$$

где $\vec{x} = (x_1, x_2, \dots, x_N)$ вектор, составленный из N отсчетов анализируемого отрезка сигнала, \mathbf{A}_r — субполосная матрица, она состоит из элементов в виде разницы синусов [4]:

$$a_r(i, k) = \begin{cases} \sin(v_r(i-k)) - \frac{\sin(v_{r-1}(i-k))}{\pi(i-k)}, & \text{при } i \neq k, \\ (v_r - v_{r-1}) / \pi, & \text{при } i = k \end{cases} \quad (12)$$

где v_r — верхняя граница r -го частотного интервала; индексы i, k меняются от 1 до N , r — это номер частотного интервала, $r = 1, 2, \dots, R$.

Разбиение на интервалы должно удовлетворять условию:

$$\begin{aligned} \Omega_r &= [-v_r, -v_{r-1}) \cup [v_{r-1}, v_r) \\ v_0 &= 0, \\ \Omega_R &= \bigcup_{r=1}^R \Omega_r = [-\pi, \pi) \end{aligned} \quad (13)$$

Квадратичная форма (11) соответствует вычислению части энергии отрезка сигнала, содержащегося в r -м частотном интервале

$$S_r(x) = \frac{1}{2\pi} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega, \quad (14)$$

где

$$X(\omega) = \sum_{n=1}^N x_n e^{-j(n-1)\omega}, \quad j = \sqrt{-1} \quad (15)$$

это трансформанта Фурье [6] с областью определения

$$-\pi \leq \omega \leq \pi. \quad (16)$$

Имеет место и обратное преобразование

$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j(n-1)\omega} d\omega. \quad (17)$$

Отсюда можно получить равенство Парсеваля для энергии сигнала в отрезке анализа

$$\|\vec{x}\|^2 = \sum_{n=1}^N x_n^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 d\omega = \sum_{r=1}^R S_r(x). \quad (18)$$

Отношение частей энергии отрезка сигнала к полной энергии отрезка сигнала определяет функцию долей энергии сигнала [4]:

$$P_r(x) = \frac{S_r(x)}{S_R(x)}, \quad (19)$$

$$S_R(x) = \sum_{r=1}^R S_r(x).$$

Так как анализируется речевой сигнал, то для нахождения распределения энергии по частотам также рассматривался мел-спектр [7]. Мел-спектральное преобразование учитывает особенности человеческого слуха, такие как критические полосы [8–10]. Коэффициенты мел-спектра вычисляются по формуле:

$$M_j = \sum_{k=1}^{\lfloor N/2 \rfloor} m_{kj} W_k, \quad (20)$$

где W_k — коэффициенты ряда Фурье для отрезка речевого сигнала, $j=1,2,\dots,K$ — это индекс фильтра с треугольной импульсной

характеристикой (рис. 1), представленной коэффициентами m_{kj} .

Коэффициенты m_{kj} вычисляются по формуле (20), здесь величины a_j — это границы критических полос слуха

$$m_{kj} = \begin{cases} \frac{k - a_{j-1}}{a_j - a_{j-1}}, & \text{если } a_{j-1} \leq k \leq a_j \\ \frac{k - a_{j+1}}{a_j - a_{j+1}}, & \text{если } a_j < k \leq a_{j+1} \\ 0, & \text{иначе.} \end{cases} \quad (21)$$

Разбиение на полосы слуха проводится равномерно в шкале мел, а затем проводится пересчёт в обычные частоты по формуле (22), здесь f — это частоты в обычной шкале, а fm — частоты в мел-шкале:

$$f = 700 \cdot \left(10^{\frac{fm}{2595}} - 1 \right) = 700 \cdot \left(e^{\frac{fm}{1127}} - 1 \right). \quad (22)$$

Для субполосного метода мел-спектр находится более естественным способом, путём разбиения оси частот сразу на критические полосы. Для этого диапазон мел частот разбивается равномерно на R интервалов:

$$fm_r = fm_{\min} + fm_{\max} \cdot r. \quad (23)$$

Здесь r — это индекс границы частотного интервала по мел шкале, $r=0,1,2,\dots,R$, fm_{\min}, fm_{\max} — начало и конец диапазона мел частот, соответствуют $f_{\min} = 0, f_{\max} = \frac{F_d}{2}$, F_d — это частота дискретизации речевого сигнала. fm_{\min}, fm_{\max} вычисляются по обратной к (22) формуле:

$$fm = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) = 127 \cdot \ln \left(1 + \frac{f}{700} \right). \quad (24)$$

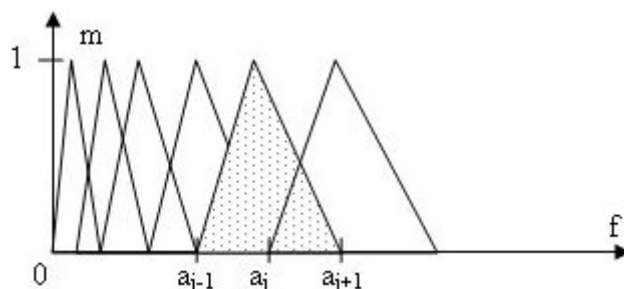


Рис. 1. Схема фильтров для вычисления мел-спектра [Fig. 1. Filter circuit for calculating the chalk spectrum]

Кроме мел-спектра, для субполосного метода применялось разбиение на полосы частот, границы которых соответствовали нотам [11], начало и конец нотного разбиения приведено в табл. 1:

Таблица 1. Соответствие нот и частот
[Table 1. Matching notes and frequencies]

Нота	Название	Частота (Hz)
A0	Ля субконтроктавы	27.50
A0#	Ля-диез субконтроктавы	29.13
B0	Си субконтроктавы	30.87
...
C8	До пятой октавы	4186.00
C8#	До-диез пятой октавы	4434.80
D8	Ре пятой октавы	4698.40

2. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для вычислительного эксперимента была взята звукозапись, оцифрованная 16 битами на частоте $F_d = 8000$ Гц. Для анализа последовательно выбирались отрезки сигнала длиной $N = 128$ отсчётов. Отрезки выбирались без перекрытия. Для каждого отрезка вычислялся спектр Фурье (9), субполосный спектр (11), мел-спектр (20) после Фурье преобразования и мел-спектр после субполосного преобразования для не равномерного разбиения шкалы частот с нотными границами и границами после пересчёта (22) из мел частот. Для мел-спектра после Фурье преобразования было взято $K = 24$ фильтра (20). Значения энергий нормировались в доли энергий (10), (19) и вычислялся объём информации (2).

На рис. 2 представлены исходный сигнал (верхний график) и информационная энтропия (нижний график), вычисленный для спектра Фурье и субполосного спектра. Видно, что графики энтропии мало отличаются друг от друга.

На рис. 3 представлены графики энтропии, вычисленные для субполосного спектра с равномерным разбиением частотной оси на полосы и для разбиения частотной оси равномерно по мел шкале с последующим переводом мел частот в обычные частоты. Из графика

видно, что значение энтропии незначительно выросло. Рассматривая вычисленную энтропию как случайную во времени величину, можно вычислить уровень значимости для гипотезы о равенстве средних значений. Величина энтропии, вычисленная для субполосного спектра с равномерным разбиением частотной оси на полосы имеет среднее значение $\mu_1 = 0.6431$ и среднеквадратичное отклонение $\sigma_1 = 0.1393$. Величина энтропии, вычисленная для субполосного спектра с разбиением частотной оси на полосы по мел шкале имеет среднее значение $\mu_2 = 0.7287$ и среднеквадратичное отклонение $\sigma_2 = 0.1134$. Количество проанализированных речевых отрезков в эксперименте было $L = 625$. Статистика для гипотезы о равенстве средних значений составляет:

$$\zeta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{L} + \frac{\sigma_2^2}{L}}} = -11.9061. \quad (25)$$

Это соответствует уровню значимости $p = 5.5 \cdot 10^{-33}$, что меньше традиционно сопоставляемого значения $p = 0.05$. Это значит, что гипотеза о равенстве средних значений не верна и различие средних значений энтропии является значимым. Разность средних значений составила 0.0855.

В отличие от разбиения частотной оси по шкале мел, разбиение с границами в нотах не однозначно повышает энтропию. На рис. 4 представлены графики информационной энтропии для субполосного спектра с равномерным разбиением оси частот и для субполосного спектра с разбиением оси частот по значениям музыкальных нот. Из графиков видно, что для некоторых отрезков речевого сигнала величина энтропии субполосного спектра с нотными границами меньше, чем для субполосного спектра с равномерными частотными полосами. Тем не менее, среднее увеличение энтропии при переходе к нотным границам произошло и составило 0.081. Величина энтропии, вычисленная для субполосного спектра с разбиением частотной оси по значениям нот имеет среднее значение $\mu_3 = 0.7241$ и среднеквадратичное отклоне-

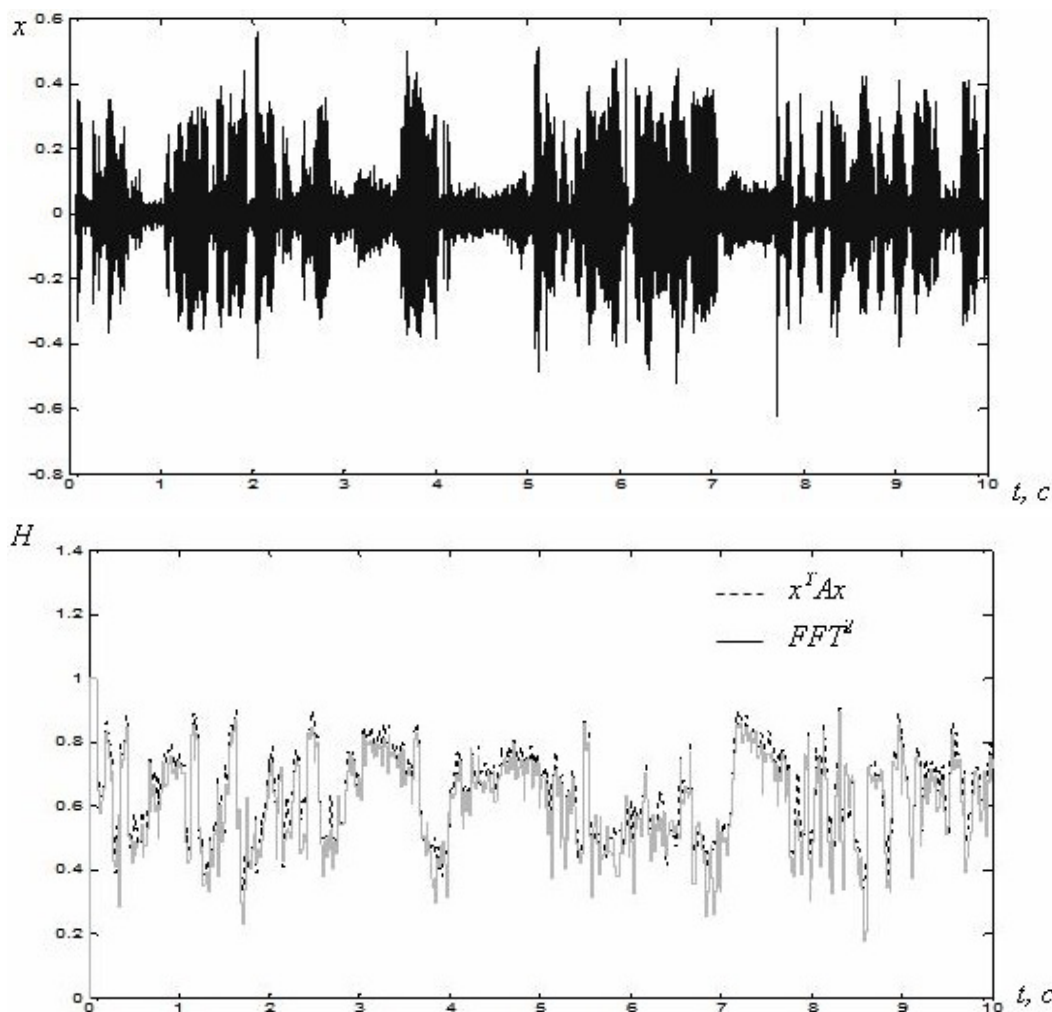


Рис. 2. Графики исходного сигнала (верхний) и информационной энтропии (нижний) для спектра Фурье и субполосного спектра [Fig. 2. Graphs of the original signal (upper) and information entropy (lower) for the Fourier spectrum and subband spectrum]

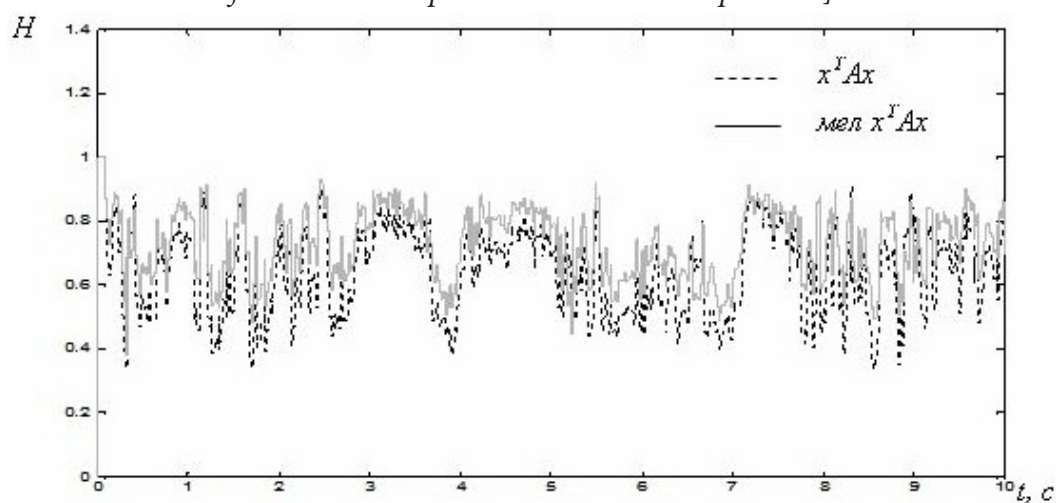


Рис. 3. Графики информационной энтропии для субполосного спектра с равномерным разбиением частотной оси и разбиением по шкале мел [Fig. 3. Graphs of information entropy for a subband spectrum with a uniform division of the frequency axis and division on a chalk scale]

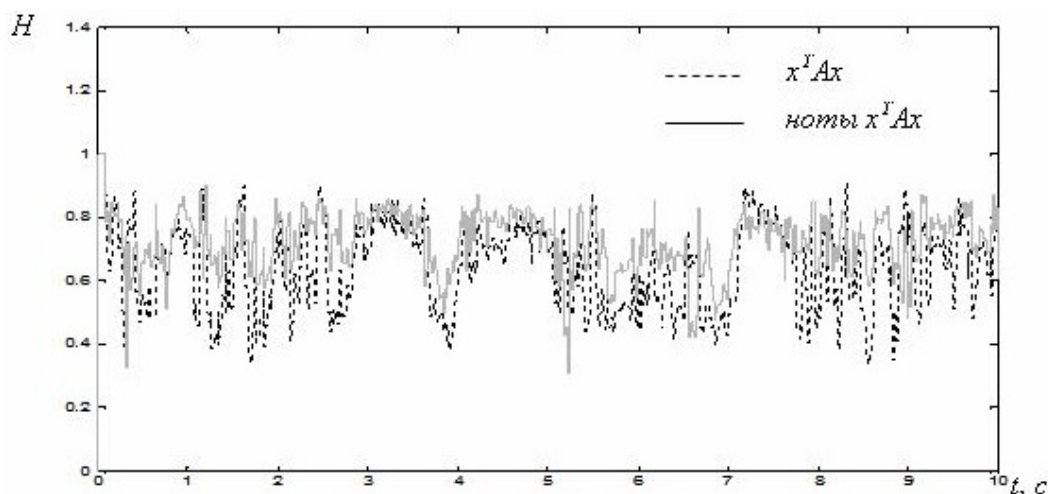


Рис. 4. Графики информационной энтропии для субполосного спектра с равномерным разбиением частотной оси и разбиением с границами по нотам
 Fig. 4. Graphs of information entropy for a subband spectrum with a uniform division of the frequency axis and division with boundaries by notes

ние $\sigma_3 = 0.0994$. Статистика для гипотезы о равенстве средних значений составляет:

$$\zeta = \frac{\mu_1 - \mu_3}{\sqrt{\frac{\sigma_1^2}{L} + \frac{\sigma_3^2}{L}}} = -11.833. \quad (26)$$

Это соответствует уровню значимости $p = 1.3 \cdot 10^{-32}$, что меньше традиционно сопоставляемого значения $p = 0.05$. Это значит, что гипотеза о равенстве средних значений не верна и различие средних значений энтропии так же, как и в предыдущем сравнении, является значимым.

Таким образом, нелинейное разбиение частотной оси на полосы для субполосного анализа повышает значение энтропии, содержащееся в получаемом энергетическом спектре. В работе было проведено ещё одно сравнение для нелинейного разбиения частотной оси. Границы частотных интервалов вычислялись по эмпирической формуле:

$$f_{test}(r) = 250 \cdot \left(1 - e^{-0.008 \cdot \frac{r}{R} \cdot f_{max}} \right) - 10 + (f_{max} - 240) \cdot \left(\frac{10}{f_{max} - 240} \right)^{\left(1 - \frac{r}{R} \right)}. \quad (27)$$

Здесь r — это индекс границы частотного интервала, $r = 0, 1, 2, \dots, R$, $f_{max} = \frac{F_d}{2}$, $F_d = 8000$ — это частота дискретизации тестового речевого сигнала.

На рис. 5 представлены графики соотношения частот в линейной, мел и тестовой шкалах. Отличие тестовой шкалы в том, что в области частот от 500 Гц до 1500 Гц разбиение на полосы происходит чаще, а в области частот от 3000 Гц до 4000 Гц, наоборот реже.

На рис. 6 представлены графики информационной энтропии для субполосного спектра с равномерным разбиением оси частот и для субполосного спектра с разбиением оси частот по эмпирической формуле.

Среднее увеличение энтропии при разбиении частотной оси по эмпирической формуле произошло и составило 0.1559, т. е. больше, чем при переходе к мел шкале или при расстановке границ по значениям музыкальных нот.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

ЗАКЛЮЧЕНИЕ

Проведенные вычислительные эксперименты показали, что, применяя субполосный метод с нелинейным разбиением частотной оси на интервалы, можно повысить информационную энтропию энергетического спектра. Приведенная эмпирическая формула не яв-

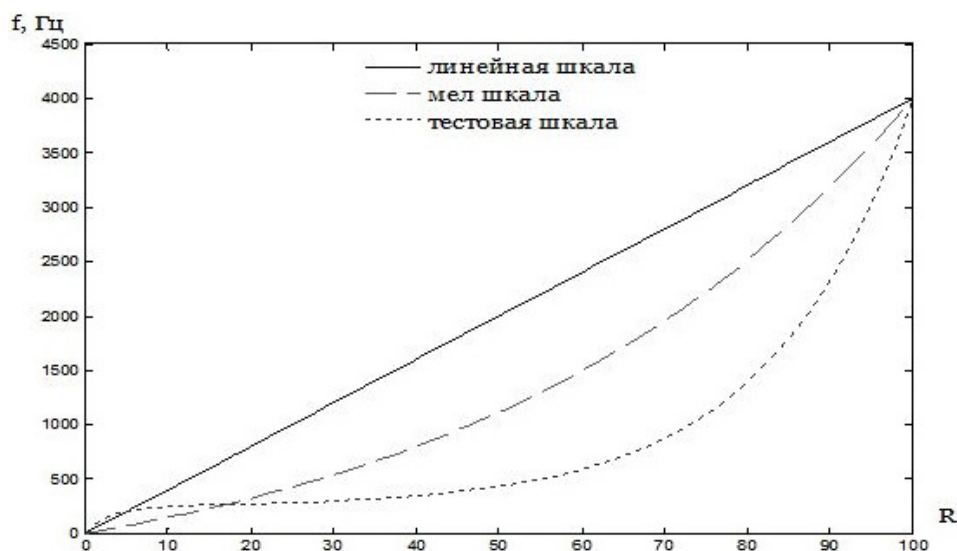


Рис. 5. Графики соотношения частот в линейной, мел и тестовой шкалах, $R=100$
 [Fig. 5. Graphs of the ratio of frequencies in linear, chalk and test scales, $R = 100$]

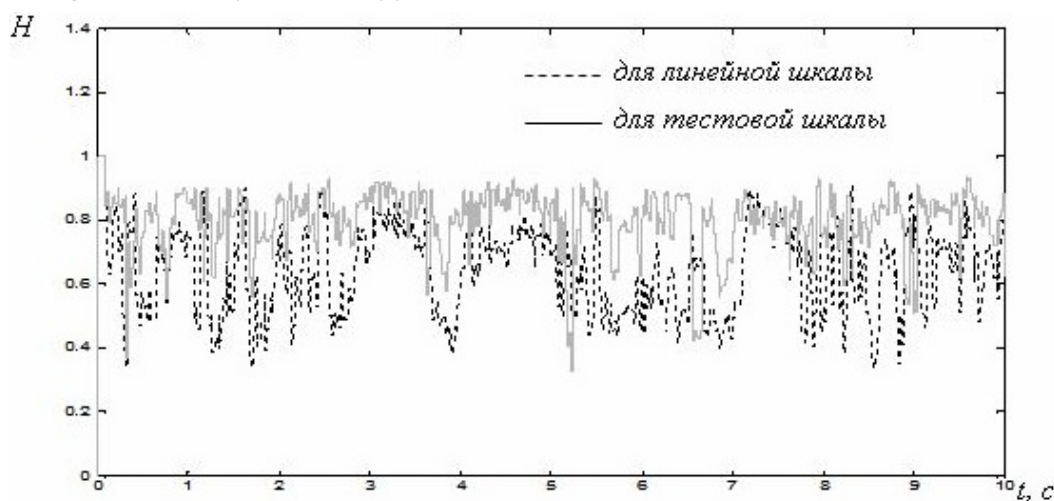


Рис. 6. Графики информационной энтропии для субполосного спектра с равномерным разбиением частотной оси и разбиением по эмпирической формуле
 [Fig. 6. Plots of Information entropy for a subband spectrum with a uniform division of the frequency axis and division by an empirical formula]

ляется оптимальной по величине энтропии, но иллюстрирует возможность повысить информативность анализа речевого сигнала за счёт учёта его особенностей.

БЛАГОДАРНОСТИ

Исследования выполнены при поддержке гранта РФФИ №№20-07-00215а

СПИСОК ЛИТЕРАТУРЫ

1. Болдышев, А. В. Параметрическое описание звуков речи в задаче распознавания / А. В. Болдышев, А. А. Медведева, Е. И. Прохоренко // Научные ведомости Белгородского государственного университета Экономика, Информатика. – 2017. – № 23 (272), вып. 44. – С. 159–169.
2. Жилияков, Е. Г. Об использовании субполосного анализа и синтеза сигналов в области определения косинус-преобразования при

решении задач сжатия речевых сигналов / Е. Г. Жилияков, Д. И. Трубицына, Е. И. Прохоренко, А. В. Болдышев // Научные ведомости Белгородского государственного университета. Экономика, Информатика. – 2019. – Т. 46, № 4, вып. 44. – С. 700–710.

3. *Shannon, C. E.* A mathematical theory of communication / С.Е. Shannon // Bell System Technical Journal. – 1948. – 27: 379-423. – P. 623–656.

4. *Жилияков, Е. Г.* Оптимальные субполосные методы анализа и синтеза сигналов конечной длительности / Е. Г. Жилияков // Автоматика и телемеханика. – 2015. – № 4. – С. 51–66.

5. *Сергиенко, А. Б.* Цифровая обработка сигналов / А. Б. Сергиенко. – Санкт-Петербург: Питер, 2005. – 604 с.

6. *Ильин, В. А.* Математический анализ. Продолжение курса / В. А. Ильин, В. А. Садовничий, Бл.Х. Сендов. – Москва : Изд-во МГУ, 1985. – 358 с.

7. *Ручай, А. Н.* Биометрическая аутентификация диктора в Matlab: учебное пособие / А. Н. Ручай. – Москва : Русайнс, 2017. – 165 с.

8. *Цвикер, Э.* Ухо как приемник информации / Э. Цвикер, Р. Фельдкеллер. – Москва : Пер. с нем. Связь, 1971. – 64 с.

9. *Алдошина, И. А.* Основы психоакустики [Электронный ресурс] / И. А. Алдошина // Книжный архив Klex.ru. – 2018. – Режим доступа: URL: <https://www.klex.ru/6n5> (дата обращения: 11 февраля 2018).

10. *Вологдин, Э. И.* Слух и восприятие звука: Учеб. пособие / Э. И. Вологдин. – СПб. СТ «Факультет ДВО», 2004. – 52 с.

11. *Бурцев, Н. Н.* О частотах нот полного звукоряда [Электронный ресурс] / Н. Н. Бурцев // Форум Обертон Про. – 2011. – Режим доступа: URL: <http://vargan.spb.ru/forum/topic/1710-бурцев-нн-о-частотах-нот-полнозвукоряда/> (дата обращения: 27.09.2017).

Жилияков Евгений Георгиевич – д-р техн. наук, проф., заведующий кафедрой информационно-телекоммуникационных систем и технологий, Белгородский государственный национальный исследовательский университет.

E-mail: zhilyakov@bsu.edu.ru

ORCID iD: <https://orcid.org/0000-0002-9456-9163>

Болдышев Алексей Владимирович — канд. техн. наук, начальник участка систем коммутаций №1 Белгородского филиала ПАО «Ростелеком».

E-mail: alexeiboldyshev@mail.ru

ORCID iD: <https://orcid.org/0000-0001-6390-6334>

Уманец Сергей Вячеславович — ведущий инженер участка систем коммутаций №1 Белгородского филиала ПАО «Ростелеком».

E-mail: 286084@bsu.edu.ru

ORCID iD: <https://orcid.org/0000-0001-5903-7586>

STUDYING THE FEATURE SPACE FOR THE DESCRIPTION OF SEGMENTS OF SPEECH SIGNALS IN SPEECH RECOGNITION PROBLEMS

© 2020 E. G. Zhilyakov¹, A. V. Boldyshev^{✉2}, S. V. Umanets²

¹Belgorod State Research University,
85, Pobeda Street, 308015 Belgorod, Russian Federation

²Belgorod branch of PJSC «Rostelekom»,
81, Bogdana Khmel'nitskogo Avenue, 308009 Belgorod, Russian Federation

Annotation. The article considers a set of features for the description of segments of speech signals with regard to their informative value. Feature sets are used in recognition problems. Effectively solving recognition problems largely depends on the information content of the feature set. The information content of a set of features is estimated using information entropy. In this article, we compare information entropy of the feature sets obtained using the Fourier method and the sub-band method. The article also considers modifications to these methods, namely the transition to the mel scale, used to analyse speech signals. In addition to the mel scale, other frequency axis distortions are considered for the sub-band method. The article presents the results of calculating the information entropy performed using different methods of obtaining a set of features describing segments of speech signals. The study demonstrated that the information content of the feature set increases when using the sub-band method with a nonlinear division of the frequency axis into analysed bands.

Keywords: speech signal, features, sub-band method, mel scale, frequency axis, fraction of energy, entropy.

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. Boldyshev A. V., Medvedeva A. A., Prokhorenko E. I. Parametric description of speech sounds in the recognition problem // Scientific Bulletin of Belgorod State University Economics, Informatics. 2017. No 23 (272). V. 44. P. 159–169.

2. Zhilyakov E. G., Boldyshev A. V., Trubitsyna D. I., Prokhorenko E. I. On the use of subband analysis and synthesis of signals in the field of determining the cosine transform when solving problems of compression of speech signals //

Scientific Bulletin of Belgorod State University Economics, Informatics. 2019. V. 46. No 4(44). P. 700–710.

3. Shannon C. E. A mathematical theory of communication // Bell System Technical Journal. 1948. (27) P. 379–423, 623–656.

4. Zhilyakov E. G. Optimal sub-band methods for analysis and synthesis of finite-duration signals // Automation and Remote Control. 2015. No 4. P. 51–66. DOI: 10.1134/S0005117915040049.

5. Sergienko A. B. Cifrovaya obrabotka signalov [Digital signal processing] St. Petersburg : Piter publ, 2005 (in Russian).

6. Il'in V. A., Sadovnichiy V. A., Sendov B. H. Matematicheskiy analiz. Prodolzenie kursa [Mathematical analysis. Continuation of the course]. Moscow : MSU publ. 1985 (in Russian).

7. Ruchai A. N. Biometricheskaya autentifikatsiya dictora v Matlab: uchebnoe posobie [Biometric Speaker Authentication in Matlab: A Tutorial] Moscow : Rusains publ. 2017 (in Russian).

✉ Boldyshev Alexey V.
e-mail: alexeboldyshev@mail.ru

8. *Tsviker E., Feldkeller E.* Ukho kak priemnik informatsii [The ear as a receiver of information] Moscow : Svyaz publ. 1971.

9. *Aldoshina I. A.* Basics of psychoacoustics. Book archive Klex.ru. 2018. Available at: <https://www.klex.ru/6n5> (accessed: 15th February 2017).

10. *Vologdin E. I.* Slukh i vospriyatie zvuka: uchebnoe posobie [Hearing and Sound Perception: A Study Guide] St. Petersburg : Facultet DVO publ. 2004 (in Russian)

11. *Burtsev N. N.* About full scale note frequencies Oberton Pro forum. 2011. Available at: <http://vargan.spb.ru/forum/topic/1710-бурцев-нн-о-частотах-нот-полногозвукоряда/> (accessed: 27th September 2017).

Zhilyakov Evgeniy G. — DSc in Technical Sciences, Professor, Head of the Department of Information and Telecommunication Systems and Technologies, Belgorod State National Research University.
E-mail: zhilyakov@bsu.edu.ru
ORCID iD: <https://orcid.org/0000-0002-9456-9163>

Boldyshev Alexey V. — PhD in Technical Sciences, Head of network switching system unit No.1 of the Belgorod branch of PAO “Rostelekom”.
E-mail: alexeiboldyshev@mail.ru
ORCID iD: <https://orcid.org/0000-0001-6390-6334>

Umanets Sergei V. — lead engineer at network switching system unit No.1 of the Belgorod branch of PAO “Rostelekom”.
E-mail: 286084@bsu.edu.ru
ORCID iD: <https://orcid.org/0000-0001-5903-7586>