

ПРИМЕНЕНИЕ АЛГОРИТМА АППРОКСИМАЦИИ ГРАФИКА ДОЛЕЙ ЭНЕРГИИ ДЛЯ ОПРЕДЕЛЕНИЯ ПАУЗ В РЕЧЕВОМ СИГНАЛЕ

© 2021 Т. Н. Балабанова¹✉, А. В. Болдышев², С. В. Уманец²

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

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

Аннотация. В данной работе рассматривается речевой сигнал как набор фрагментов, содержащих речевые компоненты и фрагменты с шумами, соответствующие паузам между словами. Ставится задача по составлению решающей функции, способной принять или отвергнуть гипотезу об отсутствии речи в отрезке речевого сигнала. На основе субполосного метода для отрезка речевого сигнала составляется его распределение энергий по частотам. Для этого распределения в дальнейшем применяется процедура аппроксимации смесью радиально-базисными функциями (функциями Гаусса). Смесью представляет собой взвешенную сумму радиально-базисных функций и равномерно-распределённой составляющей. По соотношению максимальных значений компонент смеси составляется решающее правило. Для проведения вычислительного эксперимента вводится нелинейность «зона нечувствительности», выбор которой обусловлен особенностями электрической активности путей и центров слуховой системы. В работе приводится результат применения алгоритма определения пауз в речевом сигнале. В качестве рабочего материала использовалась база размеченных речевых фрагментов американского агентства передовых оборонных исследовательских проектов DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Всего было обработано 100 звукозаписей, размер отрезка анализа был взят 9 миллисекунд, частота дискретизации 16000Гц. Для проверки работоспособности предлагаемого алгоритма были оценены ошибки первого рода «пропуск цели» — когда алгоритм не начал отмечать паузу, но такая отметка присутствует при ручной расстановке, а также ошибки второго рода «ложная тревога» — когда произошла ошибочная постановка паузы. Полученные в ходе вычислительных экспериментов результаты позволяют судить о достаточно высокой эффективности предлагаемого подхода для определения пауз в речевом сигнале.

Ключевые слова: доли энергии, смесь радиально-базисных функций, смесь Гауссовых функций, решающая функция.

ВВЕДЕНИЕ

Согласно проведенным исследованиям [1–4] одной из главных тенденций развития информационных технологий является со-

здание наиболее естественные для человека средства общения с компьютерной техникой, и ведущая роль отводится именно управлению с помощью голосового ввода информации [5]. Не менее популярным направлением также считается создание искусственного интеллекта. Алгоритмы, применяемые для автоматического распознавания речи, являются неотъемлемым этапом при создании искус-

✉ Балабанова Татьяна Николаевна
e-mail: sozonova@bsu.edu.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.

The content is available under Creative Commons Attribution 4.0 License.

ственного интеллекта [6]. Поэтому решение задачи разработки высокоточного алгоритма распознавания речи является сегодня актуальной задачей. Распознавание речи – это процесс анализа речевого сигнала, в результате которого определяется элемент устной речи, под воздействием которого сформировался анализируемый отрезок речевого сигнала. Речевой сигнал — это результат регистрации электромагнитных колебаний на выходе микрофона от воздействия акустических колебаний во время разговора.

В речевом сигнале присутствуют фрагменты, порождённые при произнесении слов (речевые компоненты), и фрагменты, содержащие шумы в паузах между словами (отсутствие речи). Одна из процедур в задачах обработки речевого сигнала — это обнаружение границы пауза/звук. Такие процедуры принято называть детекторами активности речи и именно с неё начинается первичная сегментация речевого сигнала. Часто для определения активности речи используют различия в энергиях отрезков речевого сигнала в паузах и при наличии речи. Для такого подхода необходимо проанализировать достаточно длительные отрезки речевого сигнала, что иногда бывает неприемлемо для решения конкретной задачи. Важной задачей является определение признаков, по которым будет адекватно обнаруживаться различие в свойствах шумов в паузах и речевых компонент при краткосрочном анализе.

В работе приведено описание метода обнаружения пауз, который основан на представлении долей энергии в виде смеси Гауссовых функций.

Задача состоит в том, чтобы, анализируя отрезки речевого сигнала, принять решение в пользу одной из двух гипотез. Основная (нулевая) гипотеза при определении паузы в речевом сигнале имеет формулировку:

H0: анализируемый отрезок речевого сигнала получен в отсутствие речи, т. е. порождён шумом в паузах речи.

Альтернативная гипотеза формулируется так:

H1: анализируемый отрезок речевого сигнала получен в присутствии речи.

Необходимо построить решающую функцию, позволяющую на основе обработки отрезка речевого сигнала либо принять основную гипотезу, либо отвергнуть её.

1. МЕТОДЫ ИССЛЕДОВАНИЯ

Фрагмент речевого сигнала $x(t)$ после оцифровки представляет собой вектор $\vec{x} = (x_1, x_2, \dots, x_N)^T$ (здесь T — знак транспонирования), состоящий из значений сигнала (функция времени), которые соответствуют значениям аргумента $n\Delta t$, т. е.

$$x_n = x(n\Delta t), \quad n = 1, \dots, N, \quad (1)$$

где Δt — интервал дискретизации по времени.

Для этого вектора трансформанта Фурье представляет собой функцию

$$X(\omega) = \sum_{n=1}^N x_n e^{-j(n-1)\omega}, \quad j = \sqrt{-1}. \quad (2)$$

Областью определения этой функции является нормированная частота

$$-\pi \leq \omega \leq \pi. \quad (3)$$

Имеет место и обратное преобразование

$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j(n-1)\omega} d\omega. \quad (4)$$

Отсюда можно получить равенство Парсеваля [3]

$$\|\vec{x}\|^2 = \sum_{n=1}^N x_n^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 d\omega = \sum_{r=1}^R W_r(x), \quad (5)$$

где величина

$$W_r(x) = \frac{1}{2\pi} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega \quad (6)$$

представляет собой часть энергии отрезка сигнала, содержащейся в r -м частотном интервале

$$\Omega_r = [-\omega_{2,r}, -\omega_{1,r}] \cup [\omega_{1,r}, \omega_{2,r}]. \quad (7)$$

Общее количество частотных интервалов равно R .

Распределение энергии по частотам легко вычисляется субполосным методом (здесь T — знак транспонирования)

$$W_r = \vec{x} \cdot \mathbf{A}_r \cdot \vec{x}^T, \quad (8)$$

где \mathbf{A}_r — субполосная матрица, она состоит из элементов в виде разницы синусов [7]:

$$a_r(i, k) = \begin{cases} \frac{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))}{\pi(i-k)}, & i \neq k, \\ (v_r - v_{r-1}) / \pi, & i = k \end{cases} \quad (9)$$

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:

$$\omega_{2,r} = \omega_{1,r+1}, \quad r = 1, 2, \dots, R, \quad \omega_{1,1} = 0, \quad \Omega_R = \bigcup_{r=1}^R \Omega_r = [-\pi, \pi]. \quad (10)$$

Отношение частей энергии отрезка сигнала к полной энергии отрезка сигнала определяет функцию долей энергии сигнала [8]:

$$P_r(x) = \frac{W_r(x)}{W_R(x)}, \quad W_R(x) = \sum_{r=1}^R W_r(x). \quad (11)$$

Функция долей энергии рассматривается как функция от номера частотного интервала при неизменном отрезке сигнала, поэтому аргумент x будет игнорироваться:

$$P_r = P(r). \quad (12)$$

Эта функция представляется в виде смеси:

$$\hat{P}_r = \sum_{k=1}^K \alpha_k B_{r,k} + \alpha_{K+1} C \quad (13)$$

из K радиально-базисных функций вида

$$B_{r,k} = \frac{e^{-\frac{(r-\mu_k)^2}{2\sigma_k^2}}}{\sum_{s=1}^R e^{-\frac{(s-\mu_k)^2}{2\sigma_k^2}}}; \quad k = 1, 2, \dots, K; \quad (14)$$

с параметрами

$$1 \leq \mu_k \leq R, \quad 0 < \sigma_k < \infty \quad (15)$$

и свойствами

$$B_{r,k} \geq 0, \quad r = 1, 2, \dots, R; \quad k = 1, 2, \dots, K; \quad \sum_{r=1}^R B_{r,k} = 1, \quad (16)$$

и равномерно-распределённой составляющей, представленной константой

$$C = \frac{1}{R}, \quad (17)$$

здесь K — это заданное количество компонент, α_k — весовые коэффициенты.

В работе [9] эта задача решается. В данной работе предлагается сравнить максимальные значения Гауссовых функций (14) и значения равномерно-распределённой составляющей (17), формирующих смесь для аппроксимации кривой долей энергии (12) отрезка речевого сигнала. Результатом сравнения будет индекс компоненты смеси, индекс $K+1$ соответствует равномерно-распределённой составляющей:

$$ind = \operatorname{arg} \max_k (\max(\alpha_k B_k), \alpha_{K+1} C). \quad (18)$$

Максимальное значение для Гауссовой функции будет в точке μ_k , поэтому:

$$\max(\alpha_k B_k) = \frac{\alpha_k}{\sum_{s=1}^R e^{-\frac{(s-\mu_k)^2}{2\sigma_k^2}}}. \quad (19)$$

Решающее правило следующее: если индекс компоненты $ind = K+1$, что соответствует преобладанию шума, то принимается нулевая гипотеза. Если в результате сравнения определился индекс $ind < K+1$, то принимается альтернативная гипотеза.

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 $N = 130$ отсчётов. Отрезки выбирались с перекрытием на 98 %. Количество частотных интервалов для субполосного вычисления энергий $R = 100$.

Разбиение на частотные интервалы проводилось равномерно в шкале мел [10], а затем проводился пересчёт в обычные частоты по формуле (22). Для этого диапазон мел частот разбивается равномерно на R интервалов:

$$fm_r = fm_{\min} + fm_{\max} \cdot r. \quad (20)$$

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 формуле (21).

$$\begin{aligned}
 fm &= 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) = \\
 &= 1127 \cdot \ln \left(1 + \frac{f}{700} \right), \quad (21)
 \end{aligned}$$

здесь f — частота в Герцах, fm — частота в шкале мел

Обратное преобразование из шкалы мел в шкалу Герц:

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

Для учёта порога слышимости была введена нелинейность — «зона нечувствительности». Выбор типа нелинейности обусловлен особенностями электрической активности путей и центров слуховой системы [11]. Даже при отсутствии раздражителя по волокнам слухового нерва проходят случайные импульсы с достаточно высокой частотой, около 100 в секунду. При наличии звукового раздражения частота импульсов нарастает и остаётся повышенной пропорционально интенсивности действующего звука. Возбуждение нейронов длится всё время, пока действует звук. На рис. 1 представлены графики предлагаемой нелинейности в виде кусочно-непрерывной функции и в виде гладкой аппроксимации.



Рис. 1. График нелинейности
[Fig. 1. Graph of non-linearity]

Кусочно-непрерывная функция, задающая нелинейность:

$$\hat{W} = \begin{cases} W, & W > W_{thr} \\ W_{thr}, & W \leq W_{thr} \end{cases}. \quad (23)$$

гладкая аппроксимация задавалась функцией:

$$\hat{W} = W_{thr} + \frac{W^2}{W + W_{thr}}. \quad (24)$$

Нелинейность применялась к вычисляемым значениям спектра сигнала. Спектр сигнала вычислялся субполосным методом (8). Пороговое значение $W_{thr} = 0.001$ было принято как эквивалент порога слышимости -60 дБ [12].

На рис. 2 приведен результат обработки алгоритма определения паузы в речевом сигнале. В верхней части приведен график исходного речевого сигнала и график индикатора принятия нулевой гипотезы, индикатор равен нулю при принятии нулевой гипотезы. В нижней части приведены графики изменения максимальных значений компонентов смеси $\max(\alpha_k B_k)$ и $\alpha_{k+1} C$. Количество компонент смеси было взято $K = 3$, равномерно-распределённой составляющей соответствовало $k = 4$.

Численная оценка работы алгоритма проводилась на материале из базы размеченных речевых фрагментов американского агентства передовых оборонных исследовательских проектов DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [13]. Для эксперимента были взяты 100 звукозаписей, размер отрезка анализа был взят 9 миллисекунд, что при частоте дискретизации $F_d = 16000$ Гц соответствовало длине отрезка сигнала $N = 144$ отсчёта. Отрезки выбирались с перекрытием на 98 %. Количество частотных интервалов $R = 50$, деление на частотные интервалы было неравномерное, по шкале мел (20–22). Пороговое значение $W_{thr} = 0.0001$ было принято как эквивалент порога слышимости -80 дБ.

На рис. 3 представлен один из результатов работы алгоритма, где синяя линия — это сигнал из звукозаписи, красные линии — отметки, сделанные вручную, голубые отметки получены с помощью алгоритма.

Как видно из результатов, возможно возникновение ошибочных решений о принадлежности отрезка речевого сигнала к паузе и наоборот.

Для проверки и численной оценки адекватности работы алгоритма надо подсчитать интервал времени, когда отметки ручной расстановки пауз и отметкой автоматической расстановки перекрываются:

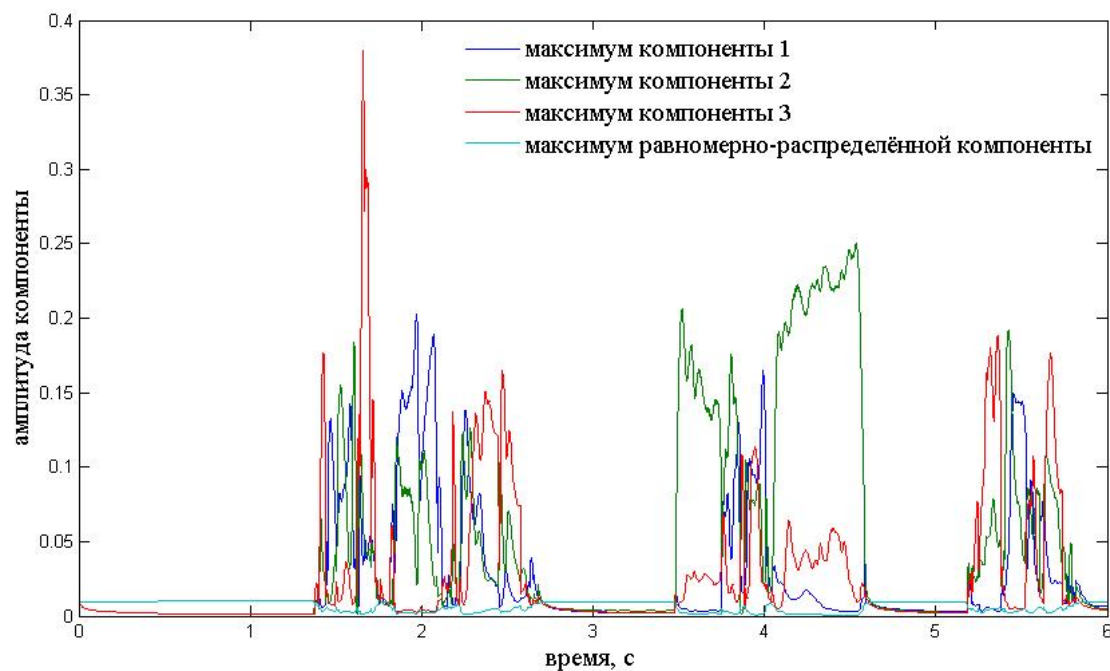


Рис. 2. Графики речевого сигнала и индикатор принятия основной гипотезы (верхний), а также графики изменения максимальных значений компонентов смеси ($K=3$) (нижний) [Fig. 2. Graphs of the speech signal and the indicator of acceptance of the main hypothesis (upper), as well as graphs of changes in the maximum values of the mixture components ($K=3$) (lower)]

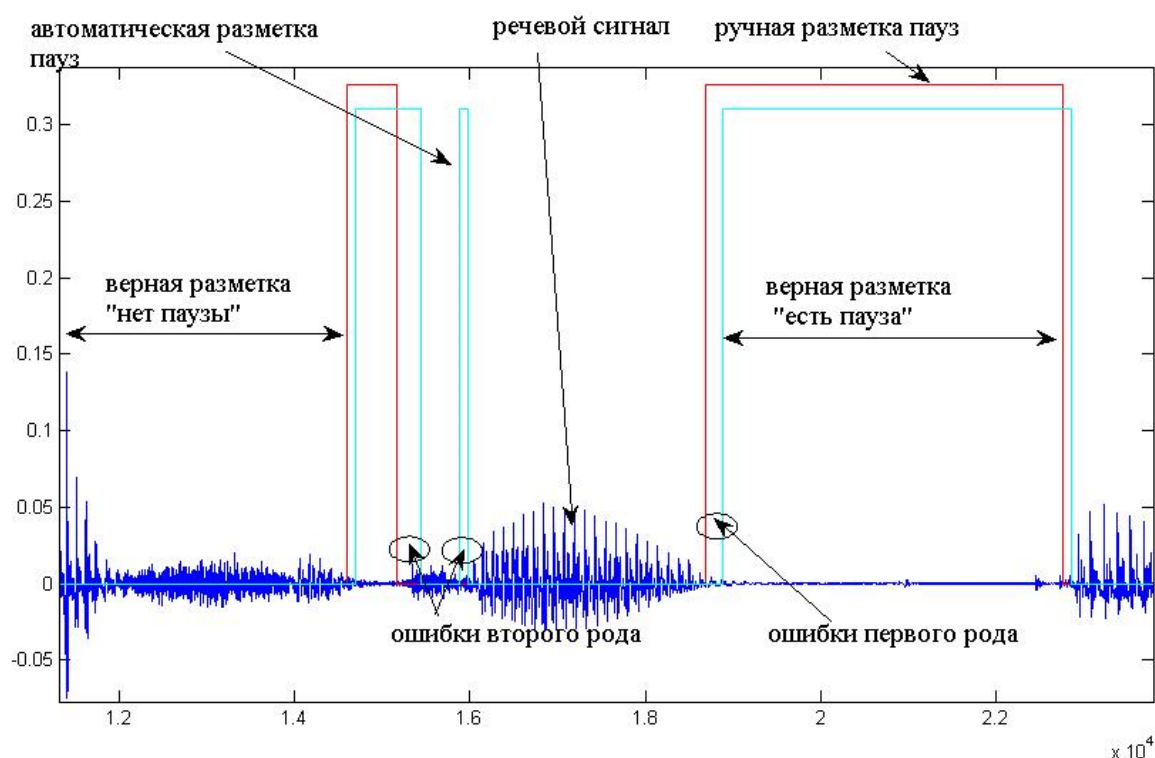


Рис. 3. Графики речевого сигнала и отметки расстановки пауз [Fig. 3. Graphs of the speech signal and pause markings]

$$C_0 = \sum_{i,j} \alpha_i \cap \beta_j + \sum_{i,j} \bar{\alpha}_i \cap \bar{\beta}_j, \quad (25)$$

где C_0 — интервал времени верных постановок паузы, $\beta \in \{B\}$ — множество интервалов ручной расстановки пауз, $\alpha \in \{A\}$ — множество интервалов автоматической расстановки пауз, $\bar{\alpha}_i, \bar{\beta}_j$ — интервалы отсутствия пауз.

Кроме количества верных оценок, ещё необходимо определить интервал времени, когда алгоритм не начал отмечать паузу, но такая отметка присутствует при ручной расстановке, это ошибки первого рода «пропуск цели».

$$C_I = \bigcup_j \beta_j \setminus C_0, \quad (26)$$

где C_I — интервал времени ошибок первого рода, как множество интервалов ручной расстановки без множества интервалов верных постановок.

Также необходимо определить интервалы времени, когда произошла ошибочная постановка паузы, это ошибки второго рода «ложная тревога».

$$C_{II} = \bigcup_j \alpha_j \setminus C_0, \quad (27)$$

где C_{II} — интервал времени ошибок второго рода, как множество интервалов автоматической расстановки без множества интервалов верных постановок.

Оценка алгоритма проводилась по критериям точности φ и полноты χ

$$\varphi = \frac{C_0}{C_0 + C_I} \quad (28)$$

$$\chi = \frac{C_0}{C_0 + C_{II}}. \quad (29)$$

По результатам эксперимента точность оказалась $\varphi = 0.932$, полнота при этом составила $\chi = 0.907$.

ЗАКЛЮЧЕНИЕ

Апробация разработанного подхода показала достаточно высокую эффективность и точность разделения анализируемых речевых сигналов на участки, порождённые при произнесении слов, т.е. содержащие речевые компоненты, и участки, содержащие шумы в паузах между словами, т.е. относящиеся к моментам отсутствия речи.

КОНФЛИКТ ИНТЕРЕСОВ

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

СПИСОК ЛИТЕРАТУРЫ

1. Алиса: голосовой помощник от компании Яндекс. Корпоративный блок компании Яндекс. – Режим доступа: <https://yandex.ru/blog/company/alisa> (10.10.2017)
2. Alexa: Amazon Unveils a Listening, Talking, Music-Playing Speaker for Your Home / Bloomberg Businessweek. Bloomberg L. P. – Электрон. журн. – Режим доступа: <https://www.bloomberg.com/news/articles/2014-11-06/amazon-echo-is-alistening-talking-music-playing-speaker-for-your-home> (Дата обращения: 06.11.2014)
3. Apple – Siri – Frequently Asked Questions. Apple Inc. – Режим доступа: <https://www.apple.com/ios/siri/> (Дата обращения: 04.10.2011)
4. Интервью с директором ООО «Речевые Технологии» Виталием Киселевым / Информационные технологии в Беларуси. – Электрон. журн. – Режим доступа: <http://it.tut.by/news/97283.html> (Дата обращения: 11.10.2015)
5. Voice Search in Russian, Polish, Czech and Turkish. – Режим доступа: <http://googlemobile.blogspot.ru/2010/10/voice-search-inrussian-polish-czech.html> (Дата обращения: 21.10.2010)
6. Bishop, C. M. Neural Networks for Pattern Recognition / C. M. Bishop. Oxford: Oxford University Press, 1995. – 498 p.
7. Ильин, В. А. Математический анализ. Продолжение курса / В. А. Ильин, В. А. Садовничий, Бл. Х. Сендов. – Москва : МГУ, 1985. – 358 с.
8. Жилияков, Е. Г. Оптимальные субполосные методы анализа и синтеза сигналов конечной длительности / Е. Г. Жилияков // Автоматика и телемеханика. – 2015. – № 4; Autom. Remote Control, 76:4.
9. Уманец, С. В. Аппроксимация долей энергии фрагмента звукового сигнала смесью радиально-базисных функций / С. В. Уманец // Научные Ведомости БелГУ. Сер. Экономика. Информатика. – 2018. – Т. 45, № 4 – С. 698–708.

10. Ручай, А. Н. Биометрическая аутентификация диктора в Matlab: учебное пособие / А. Н. Ручай. Москва : Русайнс, 2017. – 165 с.
11. Коротько, Г. Ф. Физиология человека. Учебное пособие, Том 2 / Г. Ф. Коротько, В. М. Покровский. – Москва : Медицина, 1997. – 656 с.
12. Алдошина, И. А. Основы Психоакустики [Электронный ресурс] / И. А. Алдошина // Цикл статей из журнала «Звукорежисер». – 2012. – №№ 1–10. – Режим доступа: http://nsk.jagannath.ru/users_files/books/Osnovy_psihoakustiki.pdf (Дата обращения: 17.05.2012)
13. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. – Режим доступа: <https://www.kaggle.com/mfekadu/darpa-timit-acousticphonetic-continuous-speech> (Дата обращения: 11.02.2020).

Балабанова Татьяна Николаевна — канд. техн. наук, доцент кафедры информационно-телекоммуникационных систем и технологий, Белгородский государственный национальный исследовательский университет.

E-mail: sozonova@bsu.edu.ru

ORCID iD: <https://orcid.org/0000-0003-3547-3433>

Болдышев Алексей Владимирович — канд. техн. наук, ведущий инженер станционного участка, Белгородский филиал ПАО «Ростелеком».

E-mail: alexeiboldyshev@mail.ru

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

Уманец Сергей Вячеславович — ведущий инженер станционного участка, Белгородский филиал ПАО «Ростелеком».

E-mail: 286084@bsu.edu.ru

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

DOI: <https://doi.org/10.17308/sait.2021.3/3740>

ISSN 1995-5499

Поступила в редакцию 20.07.2021

Подписана в печать 20.11.2021

APPLICATION OF THE ALGORITHM FOR APPROXIMATION OF THE GRAPHIC OF ENERGY SHARES FOR DETERMINING PAUSES IN A SPEECH SIGNAL

© 2021 T. N. Balabanova¹✉, A. V. Boldyshev², S. V. Umanets²

¹*Belgorod National Research University
85, Pobedy Street, 308015 Belgorod, Russian Federation*

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

Annotation. In this paper, a speech signal is considered as a set of fragments containing speech components and fragments with noises corresponding to pauses between words. The task is to formulate a decisive function capable of accepting or rejecting the hypothesis of the absence of speech in a segment of a speech signal. On the basis of the subband method for a segment of a speech signal, its energy distribution over frequencies is compiled. For this distribution, in what follows, a mixture approximation procedure is applied by radial basis functions (Gaussian

✉ Balabanova Tatiana N.
e-mail: sozonova@bsu.edu.ru

functions). The mixture is a weighted sum of radial basis functions and a uniformly distributed component. Based on the ratio of the maximum values of the components of the mixture, a decisive rule is drawn up. To carry out a computational experiment, the nonlinearity «dead zone» is introduced, the choice of which is due to the peculiarities of the electrical activity of the pathways and centers of the auditory system. The paper presents the result of applying the algorithm for determining pauses in a speech signal. The database of marked speech fragments of the American agency for advanced defense research projects DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus was used as a working material. In total, 100 sound recordings were processed, the size of the analysis segment was taken 9 milliseconds, the sampling rate was 16000Hz. To test the efficiency of the proposed algorithm, errors of the first kind “miss the target” were evaluated — when the algorithm did not start to mark a pause, but such a mark is present during manual placement, as well as errors of the second kind “false alarm” — when an erroneous setting of a pause occurred. The results obtained in the course of computational experiments make it possible to judge the sufficiently high efficiency of the proposed approach for determining pauses in a speech signal.

Keywords: energy fractions, mixture of radial-basis functions, mixture of Gaussian functions, decision function.

CONFLICT OF INTEREST

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

REFERENCES

1. Alice: a voice assistant from Yandex. Corporate block of Yandex. – Available at: <https://yandex.ru/blog/company/alisa> (accessed: 10th October 2017)
2. Alexa: Amazon Unveils a Listening, Talking, Music-Playing Speaker for Your Home / Bloomberg Businessweek. Bloomberg L. P. – Electron. zhurn. – Available at: <https://www.bloomberg.com/news/articles/2014-11-06/amazon-echo-is-alistening-talking-music-playing-speaker-for-your-home> (accessed: 6 November 2014)
3. Apple – Siri – Frequently Asked Questions. Apple Inc. – Available at: <https://www.apple.com/ios/siri/> (accessed 4th October 2011)
4. Interview with the Director of Speech Technologies LLC Vitaly Kiselev / Information Technologies in Belarus. – Electron. zhurn. – Available at: <http://it.tut.by/news/97283.html> (accessed 11th October 2015)
5. Voice Search in Russian, Polish, Czech and Turkish. – Available at: <http://googlemobile.blogspot.ru/2010/10/voice-search-inrussian-polish-czech.html> (accessed 21st October 2010)
6. Bishop C. M. (1995) Neural Networks for Pattern Recognition. Oxford: Oxford University Press. 498p.
7. Ilyin V. A., Sadovnichy V. A., Sendov Bl. H. (1985) Mathematical analysis. Continuation of the course. Moscow : Moscow State University. 358 p.
8. Zhilyakov E. G. (2015) Optimal subband methods of analysis and synthesis of signals of finite duration. // Automation and telemechanics. No. 4; Autom. Remote Control, 76: 4.
9. Umanets S. V. (2018) Approximation of the energy fractions of a fragment of a sound signal by a mixture of radial-basis functions. Umanets // Scientific Bulletin of BelSU. Ser. Economy. Computer science. Vol. 45, No. 4, P. 698–708.
10. Ruchay A. N. (2017) Biometric speaker authentication in Matlab: a tutorial. Moscow : Rusays. 165 p.
11. Korotko G. F., Pokrovsky V. M. (1997) Human physiology. Study guide, Volume 2. Moscow : Medicine. 656 p.
12. Aldoshina I. A. (2012) Fundamentals of Psychoacoustics [Electronic resource] // Series of articles from the magazine “Sound engineer”. No. 1–10. Available at: http://nsk.jagannath.ru/users_files/books/Osnovy_psihoakustiki.pdf (accessed 17th May 2012)
13. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Available at: <https://www.kaggle.com/mfekadu/darpa-timit-acousticphonetic-continuous-speech> 1. (accessed 11th February 2020)

Т. Н. Балабанова, А. В. Болдышев, С. В. Уманец

Balabanova Tatiana N. — PhD in Technical Sciences, associate professor of the department of information and telecommunication systems and technologies, Belgorod National Research University.
E-mail: sozonova@bsu.edu.ru
ORCID iD: <https://orcid.org/0000-0003-3547-3433>

Boldyshev Aleksei V. — PhD in Technical Sciences, Lead engineer of station department, Belgorod branch of PJSC «Rostelekom».
E-mail: alexeiboldyshev@mail.ru
ORCID iD: <https://orcid.org/0000-0001-6390-6334>

Umanets Sergei V. — Lead engineer of station department, Belgorod branch of PJSC «Rostelekom».
E-mail: 286084@bsu.edu.ru
ORCID iD: <https://orcid.org/0000-0001-5903-7586>