

УДК 534.78

ВАРИАЦИЯ ГРАНИЦ РЕЧЕВЫХ КОМАНД ДЛЯ УЛУЧШЕНИЯ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД ПО ИХ КРОССКОРРЕЛЯЦИОННЫМ ПОРТРЕТАМ

© 2013 В.Р. Крашенинников, Е.Ю. Лебедева, В.К. Капырин

Ульяновский государственный технический университет

Поступила в редакцию 2013 г.

Вероятность правильного распознавания речевых команд по некоторым их эталонам существенно снижается при ошибках определения границ распознаваемой команды. Предлагается способ уменьшения влияния этих ошибок путём варьирования оценок границ распознаваемых команд.

Ключевые слова: распознавание речевых команд, кросскорреляционные портреты, вариация границ

В настоящее время большое распространение получили речевые информационно-управляющие системы. Целью таких систем является снижение нагрузки на оператора, например, пилота. Достигается это через ведение управления объектом на естественном для человека языке, с помощью речевых команд (РК). Подобные системы должны функционировать в условиях сильных акустических помех. Одним из методов реализующих распознавание РК в условиях интенсивных шумов, является метод распознавания РК по их кросскорреляционным портретам (ККП) [1]. Суть данного метода состоит в следующем [2].

Имеется словарь, состоящий из N РК: K_1, K_2, \dots, K_N . Для каждой РК K_i в памяти хранятся два ее эталона. Каждый эталон представляет собой произнесение команды K_i одним диктором в разное время. Задача распознавания РК P является ее отнесение к одной из команд. Принадлежность РК P к классу K_i определяется в результате совмещения эталонного ККП и ККП распознаваемой РК по некоторой метрике $d = \rho(X, Y)$, определённой на парах ККП $X = \{x_{ik}\}$ и $Y = \{y_{ik}\}$. Распознаваемая РК относится к классу K_i с наименьшим расстоянием. При этом ККП $W(S1, S2)$ двух РК (сигналов) $S1$ и $S2$ представляет собой двумерное изображение, формируемое следующим образом. Обе РК $S1$ и $S2$ разбиваются на M равных частей. Каждая m -я строка ККП является последовательностью выборочных коэффициентов корреляции $r(m, k)$ m -го отрезка сигнала $S1$ и m -го отрезка сигнала $S2$, сдвинутого на k отсчетов:

$$r(m, k) = \frac{1}{L1} \sum_{j=0}^{L1-1} s1_{(m \cdot L1 + j)} s2_{(m \cdot L2 + k + j)} - \mu1_m \mu2_{m+k} \\ \sqrt{\sigma1_m^2 \sigma2_{m+k}^2}$$

$$(k = 0..K, m = 0..M - 2),$$

где $L1$ – длина m -го отрезка РК $S1$; $L2$ – длина m -го отрезка РК $S2$; $\mu1_m$ и $\mu2_m$ – выборочные средние, $\sigma1_m^2$ и $\sigma2_m^2$ – выборочные дисперсии этих отрезков.

В качестве эталонных ККП в памяти хранятся ККП $W(V_{i1}, V_{i2})$, $i=1, \dots, N$, где V_{ij} – есть j -й эталон i -й команды ($j=1, 2, i=1..N$). Для распознаваемой РК P строится N портретов $W(P, V_{i1})$, и P относится к классу K_i для которого минимально расстояние $\rho(W(P, V_{i1}), W(V_{i1}, V_{i2}))$.

Границы РК определяются с помощью некоторого алгоритма, при этом возможны ошибки – опережение или запаздывание, что отрицательно сказывается на вероятности правильного распознавания, так как от этих границ идет разбиение сигнала на отрезки и дальнейшее построение ККП. В данной работе предлагается использовать следующий способ снижения влияния неточной оценки границ путём их варьирования.

1) Находим оценки границ a и b распознаваемой РК с помощью некоторого алгоритма.

2) Рассматриваем по три варианта начала $a-t, a, a+t$ и конца $b-t, b, b+t$ команды, где $d=b-a$ – длина команды, x – коэффициент (доля от длины), определяемая опытным путём. Сдвиги начала и конца команды позволяют уменьшить влияние неточности определения границ. Получаем 9 комбинаций РК с их границами: РК($a+t, b-t$), РК($a+t, b$), РК($a+t, b+t$), РК($a, b-t$), РК(a, b), РК($a, b+t$), РК($a-t, b-t$), РК($a-t, b$), РК($a-t, b+t$). Для каждой из 9 комбинаций РК строим ККП с эталонной командой.

3) Для каждой РК K_i находим расстояния между её эталонным ККП и каждым из 9 построенных портретов. Получаем 9 расстояний, из ко-

Крашенинников Виктор Ростиславович, доктор технических наук, профессор, заведующий кафедрой «Прикладная математика и информатика». E-mail: kor@ulstu.ru
 Лебедева Екатерина Юрьевна, аспирантка.
 E-mail: katrisa@yandex.ru
 Капырин Владимир Константинович, аспирант.
 E-mail: kapirin@inbox.ru

Таблица 1. Результаты распознавания команд

t (мс)	Отношение сигнал/шум				
	5	4	3	2	1
0	90.65	85.75	82.1	71.4	54.1
10	94.65	91.35	88.4	77.45	62.6
20	96.1	89.7	89.2	81.0	62.9
30	94.15	91.9	90.1	81.5	63.65
40	94.75	90.55	88.45	81.2	63.65
50	94.25	90.5	86.6	80.3	60.45
70	90.85	86.7	81.0	71.3	58.25

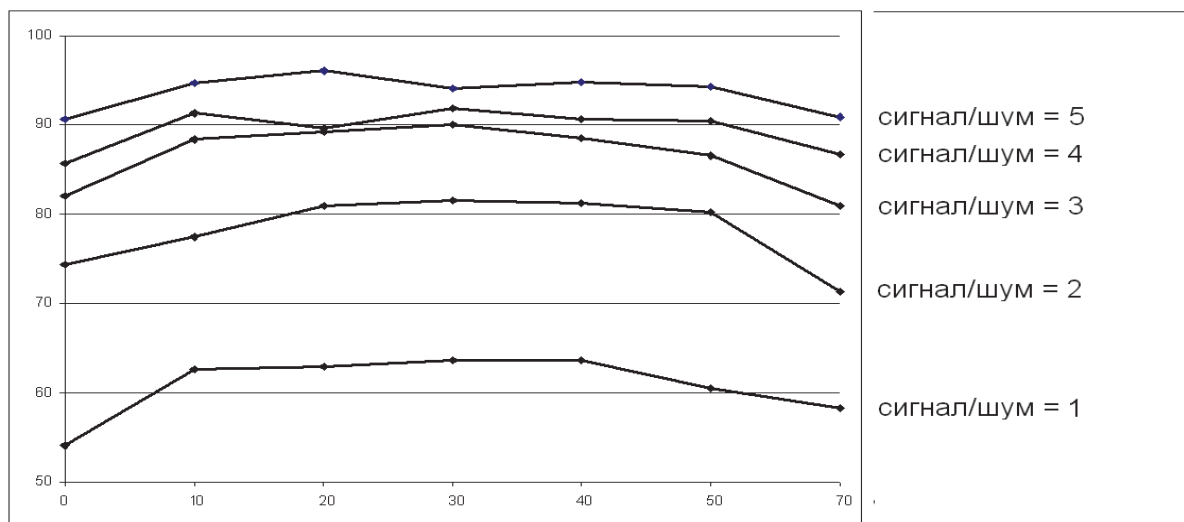


Рис. 1. Зависимость частоты правильного распознавания от величины пробного сдвига границ команд при различном отношении сигнал/шум. 0, 10, 20, 30, 40, 50, 70 мс

торых берём минимальное D_i и принимаем его за расстояние между эталонным ККП и распознаваемой РК.

4) Относим распознаваемую РК к классу K_i с минимальным расстоянием D_i .

В ходе эксперимента был использован словарь, состоящий из 20 РК авиационной тематики. В эксперименте принимали участие два диктора: мужчина и женщина. Каждая РК была произнесена по 50 раз обоими дикторами при различных отношениях сигнал/шум. В качестве сдвига t использовались временные промежутки: 0, 10, 20, 30, 40, 50, 70 мс. В табл. 1 приведён процент правильного распознавания РК при разных пробных сдвигах и отношениях сигнал/шум. Жирным шрифтом в таблице выделены наибольшие частоты правильного распознавания для каждого отношения сигнал/шум.

На рис. 1 приведены графики зависимости частоты правильного распознавания от величины сдвига при каждом отношении сигнал/шум.

Оценка значимости полученных различий в проценте правильно распознанных РК была проведена с помощью критерия Стьюдента. С уровнем значимости 0,05 была принята гипотеза о том, что вероятность правильного распознавания при варьировании границ на 20-30мс боль-

ше, чем без варьирования, то есть варьирование границ повышает вероятность правильного распознавания.

Дополнительно был проведен эксперимент, в котором в качестве пробного сдвига границ использовалась величина, равная процентной доле от длины РК: 1, 3, 5, 7, 10 и 15 %. В ходе эксперимента было получено увеличение процента правильного распознавания при использовании сдвига на 3 и 5% по сравнению с распознаванием без сдвига.

Несмотря на то, что оба способа варьирования границ показали хороший результат, логичнее использовать временной сдвиг, чем процентный, так как ошибочное определение начала и конца РК никак не зависит от длины самой команды.

Таким образом, варьирование границ с временным интервалом предлагается использовать в методе распознавания РК по их ККП для увеличения вероятности правильного распознавания.

СПИСОК ЛИТЕРАТУРЫ

1. *Krashenninikov V. R., Armer A.I., Kuznetsov V.V., Lebedeva E. Yu.* Cross-Correlation Portraits of Speech Signals in Modal-Based Speech Recognition //

Proceedings of 10th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, PRIA-8-2007. St-Petersburg, POLITECHNICA. 2010, Volume I. Pp. 105-108.

2. *Лебедева Е.Ю., Армер А.И., Ерофеев А.П.* Алгоритм

распознавания речевых команд в шумах по кросскорреляционным портретам с использованием Фурье-преобразования //Радиотехника. Математическое моделирование инфокоммуникационных систем. Вып. 175, 2012. С. 41-44.

THE IMPROVING RECOGNITION OF SPEECH COMMANDS BY CROSS-CORRELATION PORTRAITS USING BORDERS VARIATION

© 2013 V.R. Krashennnikov, E.Yu. Lebedeva, V.K. Kapirin

Ulyanovsk State Technical University

The correctness of speech commands recognition decreases if commands bounds are estimated with errors. The method to reduce errors influence using variation of recognizing commands bounds is suggested.
Key words: speech commands recognition, cross-correlation portraits, borders variation

Victor Krashennnikov, Doctor of Technics, Professor, Head at the Applied Mathematics and Informatics Department.

E-mail: kvr@ulstu.ru

Ekaterina Lebedeva. Graduate Student.

E-mail: katrisa@yandex.ru

Vladimir Kapirin, Graduate Student. E-mail: kapirin@inbox.ru