Экспериментальное исследование влияния акустических помех разных видов на результаты автоматического распознавания речевых команд

01, январь 2013

DOI: 10.7463/0113.0508614

Корсун О. Н., Яцко А. А., Финаев И. М., Чучупал В. Я.

УДК (004.934:629.7.05):001.5

Россия, МГТУ им. Н.Э. Баумана

Россия, ГосНИИАС

Россия, ВЦ им. А.А. Дородницына РАН

marmotto@rambler.ru

up1098@yandex.ru

finaevim@yandex.ru

chuchu@ccas.ru

Введение

Ключевым элементом системы речевого управления бортовым оборудованием самолетов является автоматическое распознавание речевых команд, эффективность которого определяется многими факторами, в том числе зависящими от условий полета. К числу таких факторов относятся акустические помехи различных видов. В статье представлены результаты автоматического распознавания команд, как в лабораторных условиях, так и при воздействии акустических помех различных видов, для малословарной системы, основанной на методе скрытых марковских моделей [1, 2]. Проведён анализ полученных данных. Проблема обеспечения высокой вероятности правильного автоматического распознавания речевых команд в условиях воздействия акустических помех является актуальной, ее решению в последние годы посвящено большое число исследований [3-10]. Однако в публикациях слабо представлены методики и результаты объективного количественного оценивания характеристик распознавания в условиях воздействия различных неблагоприятных факторов. Научная новизна предлагаемой статьи заключается в восполнения этого пробела в части оценивания характеристик помехоустойчивости и дикторонезависимости для систем, основанных на широко распространенном методе скрытых марковских моделей [1]. В статье предложено

дальнейшее развитие методик, разработанных для оценивания влияния пилотажных перегрузок на характеристики речи и её автоматическое распознавание [11, 12].

1 Эксперименты по оцениванию характеристик автоматического распознавания в условиях отсутствия акустических помех

Для оценивания использовалась программа, установленная на персональной ЭВМ Sony VAIO VGN-SZ6RVN, к которой подключалась авиационная гарнитура ГСП-А-189. Питание микрофона осуществлялось от батарейного источника постоянного тока напряжением 9 В через ёмкостной фильтр стандартной схемы.

Целью проводимого эксперимента являлось оценивание свойства дикторонезависимости системы распознавания, основанной на методе скрытых марковских моделей [1], описание которой дано в [2]. Для этого, все дикторы, привлекаемые к тестированию, кроме одного (Ф-в), не входили в обучающую базу данных системы. Всего обучающая база включала речевые данные 20 дикторов.

Дикторы зачитывали текст, который появлялся на экране ПЭВМ, при этом имитация нажатия гашетки речевого управления осуществлялась нажатием и отпусканием клавиши «пробел». Если во фразе хотя бы одно слово распознано ошибочно, вся фраза считалась ошибочной. Эксперимент состоял из нескольких серий. Для каждого диктора результаты эксперимента представлялись в форме, показанной в таблице 1. При этом правильно распознанные слова и фразы обозначались знаком ' + '. Обобщённые результаты приведены в таблице 2.

Таблица 1 – Результаты эксперимента для одного диктор

Фраза	Диктор Ф-в , серия № 1					Диктор Ф-в, серия № 2				
	1 слово	2 слово	3 слово	4 слово	фраза	1 слово	2 слово	3 слово	4 слово	фраза
пилотаж масштаб десять	+	+	+		+	+	+	+		+
масштаб пилотаж сто	+	+	+		+	+	+	+		+
пилотаж масштаб двести	+	+	+		+	+	+	+		+
масштаб пилотаж двести	+	+	+		+	+	+	+		+
навигация масштаб пятьдесят	+	+	+		+	+	+	+		+
навигация масштаб полторы тысячи	+	+	●	+	●	+	+	●	+	●
масштаб двадцать	+	+			+	+	+			+
масштаб пятьдесят	●	+			●	+	+			+
масштаб тысяча пятьсот	●	+	+		●	+	+	+		+
масштаб больше	+	+			+	+	+			+
масштаб меньше	+	●			●	+	+			+

Таблица 2 – Количество неправильно распознанных фраз

№ п/п	Диктор	№ серии		
		1	2	3
1	П-й	4	2	2
2	Ф-в	4	1	0
3	К-н	3	1	0
4	Б-о	4	2	4
5	З-в	2	1	0
6	М-к	4	1	0

По результатам эксперимента был сделан вывод о зависимости количества ошибок от номера серии. Вероятно, в данном случае имеет место адаптации диктора к условиям эксперимента с применением автоматической системы распознавания, что значительно снижает количество неправильно распознанных фраз. На основе таблицы 2 также был сделан вывод о том, что результаты распознавания для диктора, речевые данные которого использовались в обучающей базе (диктор Финаев), практически не отличаются от остальных. Это указывает на высокую степень дикторонезависимости при отсутствии акустических помех.

Автоматическое распознавание в лабораторных условиях проводилось с использованием того же оборудования, но по другой методике. Выполнялась запись контрольных фраз без распознавания в реальном масштабе времени. Далее выделялись участки, содержащие только одну фразу, и полученные файлы подавались на автоматическое распознавание. При выборе участков исключались дополнительные помехи (придыхание, выдох в конце фразы). Обобщенные результаты приведены в таблице 3.

Таблица 3 – Количество неправильно распознанных фраз

№ п/п	Диктор	№ серии		
		1	2	3
1	Ф-в	0	1	0
2	К-н	1	1	0
3	Б-о	0	1	0
4	З-в	0	1	0
5	М-к	1	0	1

Анализ показывает, что при данном варианте распознавания количество ошибок заметно уменьшается. Это означает, что результат распознавания в большой степени зависит от чёткости действий диктора (своевременное нажатие и отпускание гашетки, отсутствие дополнительных придыханий и других неречевых звуков).

Анализ зависимости ошибок от произносимого текста показывает, что наиболее часто ошибки распознавания имеют место только для небольшого числа фраз. Это указывает на возможность повышения качества распознавания за счёт изменения формулировки команды или путём введения иерархического распознавания.

2 Эксперименты по оцениванию характеристик автоматического распознавания при воздействии акустических помех разных видов

Целью проводимого эксперимента являлось оценивание устойчивости системы распознавания к воздействию акустических шумов. Распознавание выполнялось в двух вариантах: для малого (14 слов) и расширенного (73 слова) словарей системы распознавания.

При оценивании зависимости характеристик распознавания речи от относительного уровня помех использовались различные виды искажающего сигнала: классическая музыка, шум ненастроенного радио, шум поезда метро, фоновая речь. Данные виды помех отличаются друг от друга спектральным составом и степенью стационарности.

Искажение речевого сигнала вышеприведенными помехами производилось следующим образом. Вычислялись уровни искажаемых сигналов и помех в dB по формуле

$$L_{\text{сигнала}} \text{ (dB)} = 20 \cdot \log_{10} \hat{\sigma}_x = 10 \cdot \log_{10} \hat{D}_x.$$

Далее по полученным значениям уровней сигнала и помехи вычислялось отношение сигнал/шум

$$\text{С/Ш (dB)} = 20 \cdot \log_{10} \frac{\hat{\sigma}_x}{\hat{\sigma}_e} = 10 \cdot \log_{10} \frac{\hat{D}_x}{\hat{D}_e} = L_{\text{сигнала}} \text{ (dB)} - L_{\text{шума}} \text{ (dB)},$$

где $\hat{\sigma}_x, \hat{D}_x$ - оценки среднеквадратичного отклонения и дисперсии сигнала,

$\hat{\sigma}_e, \hat{D}_e$ - оценки среднеквадратичного отклонения и дисперсии помехи.

Вычитание из разности первоначальных уровней сигнала и помехи требуемого соотношения сигнал/шум позволяет вычислить коэффициент усиления помехи при сложении её с сигналом

$$K(\text{dB}) = -L_{\text{сигнала}}(\text{dB}) + L_{\text{шума}}(\text{dB}) - \text{требуемое С/Ш}(\text{dB}),$$

где $L_{\text{сигнала}}, L_{\text{шума}}$ - первоначальные уровни сигнала и помехи.

Умножение на данный коэффициент, а также сложение помехи с сигналом производилось в стандартном звуковом редакторе. В качестве речевого сигнала использовались фразы, произнесенные тремя разными дикторами (список из 11 фраз для каждого диктора), взятые из предыдущего теста. Для каждого диктора к фразам добавлялся аддитивный шум с различным усилением, так что были получены отношения сигнал/шум от 15 dB до 35 dB с шагом 5 dB.

Обобщенные результаты исследования представлены в таблицах 4 (малый словарь), 5 (расширенный словарь).

Таблица 4 – Итоговая таблица. Результаты распознавания команд для всех дикторов.

Количество неправильно распознанных фраз. Малый словарь

№ п/п	Диктор	Тип фонового шума, отношение сигнал/шум, dB																			
		музыка (классическая)					шум ненастроенного радио					шум в метро					другой голос				
		15	20	25	30	35	15	20	25	30	35	15	20	25	30	35	15	20	25	30	35
1	К-н	0	0	1	1	1	1	1	0	1	0	3	2	1	1	1	3	1	1	1	1
2	М-к	1	0	1	-	1	1	1	0	1	1	3	1	1	1	1	1	1	0	0	0
3	Ф-в	0	1	0	0	0	0	0	1	0	1	2	0	0	0	1	2	1	0	0	0

Таблица 5 – Итоговая таблица. Результаты распознавания команд для всех дикторов.

Количество неправильно распознанных фраз. Расширенный словарь

№ п/п	Диктор	Тип фонового шума, отношение сигнал/шум, dB																			
		музыка (классическая)					шум ненастроенного радио					шум в метро					другой голос				
		15	20	25	30	35	15	20	25	30	35	15	20	25	30	35	15	20	25	30	35
1	К-н	1	1	2	1	1	2	2	0	1	0	3	2	1	1	1	9	6	3	1	1
2	М-к	1	0	1	1	1	2	1	0	1	0	2	1	1	1	0	3	3	0	0	0
3	Ф-в	0	1	0	0	0	1	0	1	0	1	3	1	0	0	1	7	5	3	0	0

Из таблиц видно, что для помехи «классическая музыка», «шум ненастроенного радио» количество ошибок практически не зависит от соотношения сигнал/шум. Для помех типа «шум метро» количество ошибок возрастает. При увеличении отношения сигнала к шуму свыше 20 dB количество ошибок стабилизируется на достаточно низком уровне. Для помехи «другой голос», действующей в частотной полосе полезного сигнала, уровень ошибок возрастает значительно. При увеличении сигнала свыше 20 dB количество ошибок также стабилизируется на низком уровне.

В таблицах 6 и 7 показаны результаты без учета ошибок распознавания отдельных сходных по звучанию слов, а именно «тысяча» и «тысячи».

Таблица 6 – Итоговая таблица. Результаты распознавания команд для всех дикторов.

Количество неправильно распознанных фраз. Малый словарь

№ п/п	Диктор	Тип фонового шума, отношение сигнал/шум, dB																			
		музыка (классическая)					шум ненастроенного радио					шум в метро					другой голос				
		15	20	25	30	35	15	20	25	30	35	15	20	25	30	35	15	20	25	30	35
1	К-н	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	2	0	0	0	0
2	М-к	0	0	0	-	0	0	0	0	0	0	2	0	0	0	0	1	1	0	0	0
3	Ф-в	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	1	0	0	0

Таблица 7 – Итоговая таблица. Результаты распознавания команд для всех дикторов.
Количество неправильно распознанных фраз. Расширенный словарь

№ п/п	Диктор	Тип фонового шума, отношение сигнал/шум, dB																			
		музыка (классическая)					шум ненастроенного радио					шум в метро					другой голос				
		15	20	25	30	35	15	20	25	30	35	15	20	25	30	35	15	20	25	30	35
1	К-н	1	1	1	0	0	2	2	0	0	0	2	0	0	0	0	8	5	2	0	0
2	М-к	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	3	3	0	0	0
3	Ф-в	0	0	0	0	0	1	0	0	0	0	2	1	0	0	0	7	5	3	0	0

Результаты показывают, что ошибка данного вида вносит существенный вклад в конечный результат. На малом словаре для всех видов помех уровень ошибок составляет 0...1 во всех случаях, кроме помех вида «шум метро» и «другой голос», где он возрастает до 1...2. Для расширенного словаря уровень ошибок в целом составляет 0...2, а значительные ухудшения до 3...8 ошибок наблюдаются только для помехи «другой голос» при соотношениях сигнал/шум 15...20 dB.

Влияние различных видов помех легко объяснимо. В классической музыке сигналы близки к чистым тонам (пение и отдельные инструменты) и слабо коррелированы с речевыми командами. Сигнал «шум метро» представляет собой интенсивную широкополосную помеху. Наибольшее влияние на ухудшение распознавания оказывает помеха «другой голос» в силу нестационарности и, главное, вследствие коррелированности с полезным сигналом. Спектрограммы этих помех показаны на рис. 1, рис. 2.

Рисунок 1 - Спектрограмма сигнала «шум метро»

Рисунок 2 - Спектрограмма сигнала «другой голос»

Значительное влияние помех при соотношениях сигнал/шум 15...20 dB. Сравнение результатов для различных словарей показывает, что расширение словаря снижает качество распознавания при тех же видах и уровнях помех.

Заключение

Проведённые экспериментальные исследования влияния акустических помех разных видов на результаты автоматического распознавания речевых команд показали следующее.

1) В условиях отсутствия акустических помех, а также помех различных видов с соотношениями сигнал/шум свыше 20...25 dB тестируемая система автоматического распознавания речевых команд обеспечивает уровень ошибок 9...18 % после адаптации диктора к системе. Для неадаптировавшегося диктора уровень ошибок может достигать 30...40 %.

2) Для широкополосных помех и помех, коррелированных с полезным речевым сигналом («другой голос»), на уровнях 15...20 dB наблюдается существенное ухудшение качества распознавания и зависимость от размера словаря. Для малого словаря (14 слов) доля ошибок составляет 20...30 %, а для расширенного (73 слова) 30...80 %.

3) Полученные характеристики распознавания практически не зависят от индивидуальных особенностей диктора и от того, использовался ли речевой материал диктора для обучения системы.

4) Для повышения уровня безошибочного функционирования целесообразно:

- усовершенствовать логику работы системы за счет коррекции отдельных плохо распознаваемых команд;
- использовать иерархический подход к распознаванию, применяя последовательно расширенный и малый словари;
- продолжить исследования в целях разработки алгоритмов и методов помехоустойчивого распознавания речевых команд.

5) Учитывая сильную зависимость результатов распознавания от степени адаптации диктора к системе, целесообразно поставить вопрос о предъявлении к дикторам следующих требований при работе с речевым управлением:

- четкое произнесение слов;
- стремление избегать неречевых звуков (придыхание и т.д.);
- чёткая фиксация гашетки речевого управления.

Список литературы

1. Рабинер Л.Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: Обзор // ТИИЭР (Труды института инженеров по электротехнике и радиоэлектронике) : пер. с англ. 1989. Т. 77, № 2. С. 86-120. [Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. Feb. 1989. Vol. 77, no. 2. P. 257-286. DOI: 10.1109/5.18626].

2. Бондарос Ю.Г., Маковкин К.А., Чучупал В.Я. Система распознавания команд речевого интерфейса пилота для интегрированной модульной авионики // Вестник компьютерных и информационных технологий. 2007. № 4. С. 2-13.

3. Chen B., Loizou P., A Laplacian-based MMSE estimator for speech enhancement // Speech communication. 2007. Vol. 49, no. 2. P. 134-143. DOI: [10.1016/j.specom.2006.12.005](https://doi.org/10.1016/j.specom.2006.12.005)

4. Cifani S., Principi E., Rocchi C., Squartini S., Piazza F. A multichannel noise reduction front-end based on psychoacoustics for robust speech recognition in highly noisy environments // Proc. of the 2008 IEEE Hands-Free Speech Communication and Microphone Arrays (HSCMA 2008). 2008. P. 172-175. DOI: <http://dx.doi.org/10.1109/HSCMA.2008.4538714>

5. Cohen I. Relative transfer function identification using speech signals // IEEE Transactions on Speech and Audio Processing. 2004. Vol. 12, no. 5. P. 451-459. DOI: [10.1109/TSA.2004.832975](https://doi.org/10.1109/TSA.2004.832975)

6. Gales M., Young, S. An improved approach to the hidden Markov model decomposition of speech and noise // Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-92). 1992. Vol. 1. P. 233-236.

7. Gannot S., Burshtein D., Weinstein E. Signal enhancement using beamforming and nonstationarity with applications to speech // IEEE Transactions on Signal Processing. 2001. Vol. 49, no. 8. P. 1614-1626.

8. Gannot S., Cohen I. Speech enhancement based on the general transfer function and postfiltering // IEEE Transactions on Speech and Audio Processing. 2004. Vol. 12, no. 6. P. 561-571. DOI: [10.1109/TSA.2004.834599](https://doi.org/10.1109/TSA.2004.834599)

9. Garcia L., Gemello R., Mana F., Segura J. Progressive memory-based parametric non-linear feature equalization // 10th Annual Conference of the International Speech Communication Association (INTERSPEECH) (Brighton, UK, 6-10 September 2009). 2009. P. 40-43.

10. Garcia L., Segura J., Ramirez J., De La Torre A., Benitez C. Parametric nonlinear feature equalization for robust speech recognition // Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2006. Vol. 1. P. 2-12.

11. Методика экспериментального исследования влияния пилотажной перегрузки на характеристики речи в целях создания речевого интерфейса бортового оборудования самолетов / О.Н. Корсун, А.И. Иванов, В.Н. Филатов, И.В. Красавин, В.Я. Чучупал // Вестник компьютерных и информационных технологий. 2012. № 5. С. 3-7.

12. Результаты экспериментальных исследований влияния пилотажной перегрузки на характеристики речи / О.Н. Корсун, А.И. Иванов, В.Н. Филатов, И.В. Красавин, В.Я. Чучупал // Вестник компьютерных и информационных технологий. 2012. № 6. С. 3-7.

Experimental research of influence of acoustic noises of different types on results of automatic speech recognition

01, January 2013

DOI: 10.7463/0113.0508614

Korsun O.N., Yacko A.A., Finaev I.M., Chuchupal V.Ya.

Russia, Bauman Moscow State Technical University

Russia, State Research Institute of Aviation Systems

Dorodnicyn Computing Centre of RAS

marmotto@rambler.ru

up1098@yandex.ru

finaevim@yandex.ru

chuchu@ccas.ru

The article presents estimation results of automatic speech recognition by the system based on the method of hidden Markov models. The research was carried out in the following conditions: in a laboratory in the absence of acoustic noises, under the influence of acoustic noises of different types. The presented results give objective quantitative estimation of such characteristics as speaker-independence and noise-immunity of speech recognition systems with little vocabulary, based on Markov models; recommendations on improving speech recognition characteristics are formulated. The authors propose further development of the techniques developed for estimation of influence of flight overloads on speech characteristics and its automatic recognition.

Publications with keywords: [automatic speech recognition](#), [direct voice control of aircraft systems](#), [pilot interface for the aircraft cockpit](#), [robustness of voice control system to noise](#)

Publications with words: [automatic speech recognition](#), [direct voice control of aircraft systems](#), [pilot interface for the aircraft cockpit](#), [robustness of voice control system to noise](#)

References

1. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Feb. 1989, vol. 77, no. 2, pp. 257-286. DOI: 10.1109/5.18626 (Russ. ed.: Rabiner L.R. Skrytye markovskie modeli i ikh primeneniye v izbrannykh prilozheniiakh pri raspoznavanii rechi: Obzor. *TIIEP*, 1989, vol. 77, no. 2, pp. 86-120.).

2. Bondaros Iu.G., Makovkin K.A., Chuchupal V.Ia. Sistema raspoznavaniia komand rechevogo interfeisa pilota dlia integrirovannoi modul'noi avioniki [System of recognition of commands

of speech interface of pilot for integrated modular avionics]. *Vestnik komp'uternykh i informatsionnykh tekhnologii* [Herald of Computer and Information Technology], 2007, no. 4, pp. 2-13.

3. Chen B., Loizou P., A Laplacian-based MMSE estimator for speech enhancement. *Speech communication*, 2007, vol. 49, no. 2, pp. 134-143. DOI: [10.1016/j.specom.2006.12.005](https://doi.org/10.1016/j.specom.2006.12.005)

4. Cifani S., Principi E., Rocchi C., Squartini S., Piazza F. A multichannel noise reduction front-end based on psychoacoustics for robust speech recognition in highly noisy environments. *Proc. of the 2008 IEEE Hands-Free Speech Communication and Microphone Arrays (HSCMA 2008)*, 2008, pp. 172-175. DOI: <http://dx.doi.org/10.1109/HSCMA.2008.4538714>

5. Cohen I. Relative transfer function identification using speech signals. *IEEE Transactions on Speech and Audio Processing*, 2004, vol. 12, no. 5, pp. 451-459. DOI: [10.1109/TSA.2004.832975](https://doi.org/10.1109/TSA.2004.832975)

6. Gales M., Young, S. An improved approach to the hidden Markov model decomposition of speech and noise. *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-92)*, 1992, vol. 1, pp. 233-236.

7. Gannot S., Burshtein D., Weinstein E. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 2001, vol. 49, no. 8, pp. 1614-1626.

8. Gannot S., Cohen I. Speech enhancement based on the general transfer function and postfiltering. *IEEE Transactions on Speech and Audio Processing*, 2004, vol. 12, no. 6, pp. 561-571. DOI: [10.1109/TSA.2004.834599](https://doi.org/10.1109/TSA.2004.834599)

9. Garcia L., Gemello R., Mana F., Segura J. Progressive memory-based parametric non-linear feature equalization. *Proc. of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Brighton, UK, 6-10 September, 2009, pp. 40-43.

10. Garcia L., Segura J., Ramirez J., De La Torre A., Benitez C. Parametric nonlinear feature equalization for robust speech recognition. *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006, vol. 1, pp. 2-12.

11. Korsun O.N., Ivanov A.I., Filatov V.N., Krasavin I.V., Chuchupal V.Ia. Metodika eksperimental'nogo issledovaniia vliianiia pilotazhnoi peregruzki na kharakteristiki rechi v tseliakh sozdaniia rechevogo interfeisa bortovogo oborudovaniia samoletov [The technique for methodic for experimental research of overload influence on speech characteristics for avionics speech interface design]. *Vestnik komp'uternykh i informatsionnykh tekhnologii* [Herald of Computer and Information Technology], 2012, no. 5, pp. 3-7.

12. Korsun O.N., Ivanov A.I., Filatov V.N., Krasavin I.V., Chuchupal V.Ia. Rezul'taty eksperimental'nykh issledovaniy vliianiia pilotazhnoi peregruzki na kharakteristiki rechi [Experimental research of overload influence on speech characteristics]. *Vestnik komp'uternykh i informatsionnykh tekhnologii* [Herald of Computer and Information Technology], 2012, no. 6, pp. 3-7.