

Раздел II. Акустические методы и приборы в медико-биологической практике

УДК 629.7.05

Ю.Г. Бондарос, А.И. Иванов, А.А. Шишов

АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ ОПЕРАТОРОВ КРИТИЧНЫХ ПО БЕЗОПАСНОСТИ СИСТЕМ: ГИПОБАРИЯ*

Успехи в технологиях распознавания речи за последние годы привели к увеличению интереса разработчиков систем управления к речевым интерфейсам. Речевой интерфейс – это альтернативный канал обеспечения диалога оператора и управляемой системы (например, водителя и автомобиля, пилота и самолёта).

К критичным по безопасности системам можно отнести, например, все виды транспорта. Интенсификация наземного и воздушного движения, расширение функций, возлагаемых на операторов – водителей и пилотов, приводят к увеличенной рабочей нагрузке. Прямой речевой ввод команд управления критичными по безопасности системами, при надежной работе, может уменьшать рабочую нагрузку оператора, повышать безопасность и улучшать понимание ситуации.

Особенностью условий работы системы распознавания речевых команд большинства критичных по безопасности систем является широкий диапазон изменения внешних условий, в которых находится оператор-диктор. Данная работа посвящена исследованию изменчивости речи диктора-оператора в результате быстрого изменения давления воздуха окружающей среды.

Применительно к выполнению авиационных полетов диапазон изменения барометрического давления, воздействующего на оператора-диктора, зависит от динамики давления в герметичной (разгерметизированной) кабине. В данной работе этот диапазон ограничен давлением, эквивалентным давлению на высоте 4 км, что является наиболее характерным для условий гермокабины отдельных типов летательных аппаратов [1].

Аудиозаписи речи операторов выполнялись в барокамере СБК-80 ГосНИИИ ВМ МО РФ с помощью профессионального цифрового диктофона с внешним электретным микрофоном ЕК-3027 фирмы Knowles. Формат аудиозаписей MS wav, режим “моно” с квантованием по амплитуде 16 бит и квантованием по времени 32 кГц [2].

В качестве дикторов в эксперименте участвовали два испытуемого (мужчины) в возрасте 23-х и 54-х лет. Каждый диктор подвергался испытанию в одну сессию, которая состояла из 6-ти последовательно выполняемых этапов: диктование в нормальных барометрических условиях (высота – 0 км), в условиях барометрического давления, эквивалентного высотам соответственно 1 км, 2 км, 3 км, 4 км, и затем снова в наземных условиях (высота +0 км). Время изменения давления при переходе от одного этапа к другому составляло в среднем 50–60 с, время экспозиции на каждой из высот составляло 4–5 мин. Испытуемые получали истинную информацию о фактическом давлении окружающего воздуха в барокамере. Время диктования полного набора фраз теста составляло примерно 2 – 2,5 мин.

* Работа выполнена при поддержке РФФИ. Грант РФФИ 06-08-01534-а

Использовался только режим асинхронного диктования, когда диктор читает текст и таким образом сам выбирает темп диктования (в отличие от режима синхронного диктования, когда диктор произносит фразы текста по периодически возникающей на экране подсказке и следует предложенному компьютером темпу). В качестве текстового материала использованы фразы и команды из приложения Д100 ГОСТ Р 50840-95 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости», содержащего 50 фраз.

Для оценки средней энергии речевого сигнала диктора при фиксированном барометрическом давлении по множеству фраз теста вычислялось среднее $M_E(H)$ и СКО $\sigma_E(H)$ значений средней энергии речевого сигнала для каждой фразы. В табл. 1 приведены относительные данные $M_E(H)/M_E(-0)$ и $\sigma_E(H)/M_E(H)$ энергии речевых сигналов.

Таблица 1

Энергетические характеристики речевого сигнала в зависимости от высоты подъема в барокамере

Высота Н, км		-0	1	2	3	4	+0
Диктор 1	$M_E(H)/M_E(-0)$	1,0	1,64	2,16	1,5	2,35	0,93
Диктор 2	$M_E(H)/M_E(-0)$	1,0	1,2	1,3	1,0	-	-
Диктор 1	$\sigma_E(H)/M_E(H)$	0,635	0,88	0,89	0,95	0,87	0,7
Диктор 2	$\sigma_E(H)/M_E(H)$	0,42	0,33	0,357	0,37	-	-

В качестве временных характеристик речи использовались следующие параметры:

- относительная длительность произнесения теста $T_\Sigma(H)/T_\Sigma(-0)$;
- математическое ожидание M_τ и СКО σ_τ нормированной (деленной на число гласных звуков в фразе) длительности произнесения фразы;
- отношение суммарной длительности речевых отрезков $T_{\Sigma S}(H)$ теста к полной длительности $T_\Sigma(H)$ теста: $T_{\Sigma S}(H)/T_\Sigma(H)$.

Таблица 2

Временные характеристики речи в зависимости от высоты подъема в барокамере

Высота Н, км		-0	1	2	3	4	+0
Диктор 1	$T_\Sigma(H)/T_\Sigma(-0)$	1,0	1,1	1,14	1,26	1,18	1,12
Диктор 2	$T_\Sigma(H)/T_\Sigma(-0)$	1,0	0,95	0,97	0,9	-	-
Диктор 1	$M_\tau(H)/M_\tau(-0)$	1,0	1,04	1,06	1,12	1,01	1,05
Диктор 2	$M_\tau(H)/M_\tau(-0)$	1,0	1,01	1,05	1,0	-	-
Диктор 1	$\sigma_\tau(H)/M_\tau(H)$	0,136	0,162	0,159	0,203	0,139	0,173
Диктор 2	$\sigma_\tau(H)/M_\tau(H)$	0,195	0,225	0,209	0,203	-	-
Диктор 1	$T_{\Sigma S}(H)/T_\Sigma(H)$	0,83	0,76	0,78	0,78	0,65	0,76
Диктор 2	$T_{\Sigma S}(H)/T_\Sigma(H)$	0,71	0,74	0,76	0,78	-	-

Раздел II. Медицинские акустические приборы

Оценка спектра речевого сигнала диктора при фиксированном барометрическом давлении по множеству фраз теста вычислялась как среднее значение M_{F_c} и СКО σ_{F_c} по множеству средних значений энергетического спектра речевого сигнала для каждой фразы.

Таблица 3

Зависимость средней частоты спектра речи от высоты подъема в барокамере

Высота Н, км		-0	1	2	3	4	+0
Диктор 1	$M_{F_c}(H) / M_{F_c}(-0)$	1,00	0,96	0,99	1,01	1,07	1,02
Диктор 2	$M_{F_c}(H) / M_{F_c}(-0)$	1,00	1,10	1,14	1,11		
Диктор 1	$\sigma_{F_c}(H) / M_{F_c}(H)$	0,083	0,066	0,074	0,080	0,066	0,129
Диктор 2	$\sigma_{F_c}(H) / M_{F_c}(H)$	0,078	0,069	0,096	0,121	-	-

На рис. 1 и 2 приведены спектрограммы речи «экипажу разрешаю взлёт» дикторов 1 и 2. Для спектрограмм обоих дикторов характерны заметные изменения спектрального портрета речи в области высоких частот в зависимости от высоты подъема в барокамере и изменения барометрического давления. Причем выявленные изменения наиболее характерны на спектрограмме диктора 2.

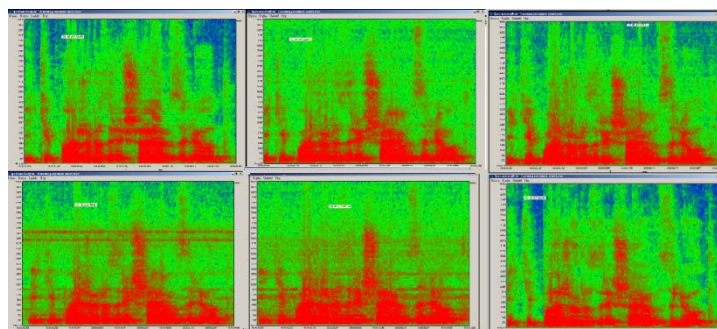


Рис. 1. Спектрограммы речи «экипажу разрешаю взлёт» диктора 1 ($H = -0, 1, 2, 3, 4, +0$ км)

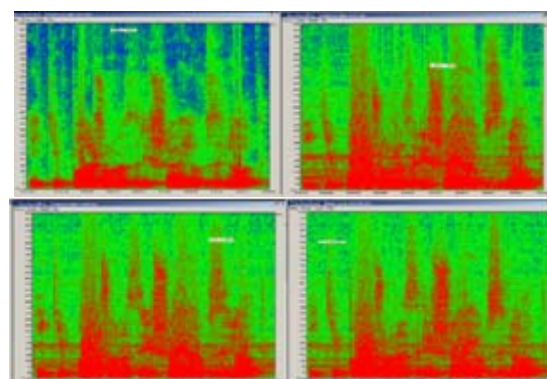


Рис. 2. Спектрограммы речи «экипажу разрешаю взлёт» диктора 2 ($H = -0, 1, 2, 3$ км)

Выводы.

1. С понижением давления воздуха изменение характеристик речевого сигнала диктора зависит от подготовленности испытуемого к воздействию гипоксической гипоксии и, возможно, от его индивидуальной высотной устойчивости.

2. Для диктора 1 с полным отсутствием опыта испытательской работы и менее высокой, чем у диктора 2 индивидуальной высотной устойчивостью с увеличением высоты наблюдается рост энергии сигнала, рост длительности произнесения теста, рост нормированной длительности произнесения фраз, уменьшение отношения суммарной длительности речевых отрезков к полной длительности теста. Для диктора 2 с увеличением высоты наблюдается неубывающее изменение энергии сигнала, уменьшение длительности произнесения теста, неубывающее изменение длительности произнесения фраз, увеличение отношения суммарной длительности речевых отрезков к полной длительности теста.

3. Для обоих дикторов с увеличением высоты наблюдается смещение центра тяжести энергетического спектра в сторону более высоких частот. Причем для диктора 2 с большей подготовленностью к переносимости экстремальных факторов авиационного полета и более высокой индивидуальной высотной устойчивостью эта закономерность более выражена.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Начала авиационной и космической медицины: Учебное пособие / Под ред. И.Б. Ушакова. – М.: Медицина для всех, 2007. – 400 с.
2. Бондарос Ю.Г., Иванов А.И., Шишов А.А. База аудиоданных “Вертикаль”, ФГУ ГосНИИ ВМ Минобороны России. М., 2008.

УДК 629.7.05

Ю.С. Перервенко, В.А. Черемушкин**ФОРМИРОВАНИЕ БАЗЫ АУДИОЗАПИСЕЙ ДИСПЕТЧЕРОВ
РОСТОВСКОГО УЗЛА СКЖД В ОПЕРАТИВНОЙ ОБСТАНОВКЕ**

Проблема контроля эмоционального и психофизиологического состояния человека-оператора является актуальной в современной ситуации возрастающей нагрузки на организм и стрессовых ситуаций. Аудиоканал (попросту речь) является естественным источником информации о состоянии человека.

Представлен алгоритм формирования базы аудиозаписей для дальнейшего анализа и обработки. Аудиозаписи получены с помощью оцифровки видеокассет, содержащих записи рабочих моментов диспетчеров ростовского узла СКЖД. Исходный сигнал содержит множество помех, определяемых акустическим несовершенством помещения, артефактами положения микрофона и др. Его аудиограмма показана на рис. 1,а.

Далее сигнал подвергнулся обработке фильтром нижних частот Блэкмана с частотой среза 4000 Гц; затем ко всему сигналу был применен программный фильтр от шума (использовался профиль шума, выделенный из сигнала), также проводилось удаление из сигнала звуков, резко выделяющихся из общего фона (технология Click/Pop Eliminator, которая автоматически определяет аномальные значения и сглаживает их, что лучше, чем вырезать эти аномалии из сигнала). Окончательный результат представлен на рис. 1,б.