

ИДЕНТИФИКАЦИЯ ЛЕТЧИКА В СОСТАВЕ ЭКИПАЖА НА ОСНОВЕ РЕЧЕВОЙ ПЕРЕДАТОЧНОЙ ФУНКЦИИ

О.Н. Корсун¹

marmotto@rambler.ru

Е.И. Михайлов²

mihevkonig@mail.ru

¹ ФГУП «ГосНИИАС», Москва, Российская Федерация² МФТИ, Долгопрудный, Московская обл., Российская Федерация

Аннотация

Рассмотрена проблема идентификации летчика в составе экипажа по голосу, что является одним из направлений совершенствования интерфейса кабины пилота современного самолета. Предложена методика идентификации личности говорящего, основанная на применении речевой передаточной функции диктора по частоте и алгоритма классификации данных k ближайших соседей. В силу особенностей задачи идентификация проведена для небольших групп операторов численностью до четырех человек. Приведены основные результаты апробации предложенной методики на экспериментальных речевых данных. Показана возможность идентификации оператора по небольшому числу кодовых слов с точностью около 97...99 % при использовании речевой передаточной функции диктора по частоте. Представлено сравнение результатов применения методики для летчиков вертолетной авиации с диагностированными заболеваниями органов слуха и для группы операторов без заболеваний органов слуха

Ключевые слова

Идентификация, речевая передаточная функция, частота, классификация, метод k ближайших соседей

Поступила 10.04.2019

© Автор(ы), 2019

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект РФФИ № 18-08-01142-а)

Введение. В последнее время наблюдается активное развитие биометрических технологий в задачах верификации и идентификации личности. Одно из перспективных направлений — использование голосовых команд [1, 2]. Голосовая идентификация обладает многими преимуществами по сравнению как с традиционными методами идентификации, так и с другими биометрическими технологиями. Так, реализовать запись голосовых сообщений гораздо проще, чем, например, провести снятие от-

печатков пальцев или просканировать сетчатку глаза. При этом задача идентификации предъявляет высокие требования к качеству алгоритмов и вероятности правильного распознавания личности. Что касается речевых технологий, достижение достаточной точности, как правило, требует серьезных вычислительных мощностей. Это, по понятным причинам, является недостатком при интеграции подобных систем в интерфейс кабины пилота современного самолета.

Использование голосовой идентификации в задаче совершенствования интерфейса кабины пилота современного самолета [3–6] мотивировано в первую очередь желанием уменьшить рабочую нагрузку на летчиков. При этом необходимость определения личности произносящего речевую команду может заключаться в выдаче доступа к управлению критически важным бортовым оборудованием, верификации в системе голосового управления летательным аппаратом или при расследовании авиационных происшествий. Таким образом, в настоящее время задача голосовой идентификации [7, 8] членов экипажа — одно из актуальных направлений развития бортовых речевых технологий. При этом используемые для решения этой задачи методики и алгоритмы должны быть достаточно простыми и не предъявлять существенных требований к вычислительным мощностям бортового оборудования.

Понятие речевой передаточной функции диктора по частоте. Наиболее значимые изменения речевого сигнала происходят в области частоты и времени. Мощность речевого сигнала в разных частотных диапазонах может отличаться на десятки децибел, что затрудняет анализ изменений в абсолютных величинах. Поэтому при анализе изменений характеристик речи целесообразно использовать приращения сигнала относительно, например, некоторого сигнала от эталонного диктора. В связи с этим предложен инструмент речевой передаточной функции диктора по частоте. Это понятие введено в работах [9, 10].

Для определения речевой передаточной функции диктора по частоте используем известное в теории автоматического управления понятие передаточной функции $W(p)$, определяемой как отношение Лаплас-изображений выходного и входного сигналов при нулевых начальных условиях. Применяя к переменной Лапласа подстановку $p = j2\pi f$ (j — мнимая единица), получаем передаточную функцию $W(f)$ от частоты f . Функция $W(f)$ содержит информацию об амплитуде и фазе изменения сигнала на частоте f . В речевом сигнале человек воспринимает его амплитуду, а фазу не учитывает. Поэтому далее рассмотрим только модуль $|W(f)|$, который и назовем речевой передаточной функцией диктора

по частоте. По теореме Винера — Хинчина для модуля передаточной функции выполняется соотношение

$$|W(f)| = \sqrt{\frac{S_y(f)}{S_x(f)}},$$

где $S_y(f)$, $S_x(f)$ — спектральные плотности выходного и входного сигналов.

Для дальнейшей работы с входным речевым сигналом используем алгоритм частотно-временного квантования, или параметризации, речевого сигнала [11]. Напомним, что на выходе получается матрица параметрического портрета слова:

$$X = \{x_{ij}\}, \quad i = 1, \dots, N_f; \quad j = 1, \dots, N_t,$$

где столбцы соответствуют N_t интервалам квантования по времени, а строки содержат значения $i = 1, \dots, N_f$ частотных составляющих. Число частотных полос задается в пределах 30–40, длительность интервала квантования по времени 10...25 мс. Применим алгоритм параметризации речевого сигнала к N реализациям каждого слова и вычислим средний параметрический портрет:

$$\bar{X} = \frac{1}{N} \sum_{k=1}^N \{x_{ij}\}_k, \quad i = 1, \dots, N_f; \quad j = 1, \dots, N_t.$$

Поскольку интересуют изменения только частотных свойств речевого сигнала, выполним осреднение матрицы \bar{X} по временным кадрам. Тогда получим вектор a средних амплитуд частотных составляющих размерности N_f , принадлежащих i -й частотной полосе:

$$a = \frac{1}{N_t} \sum_{j=1}^{N_t} \{\bar{x}_{ij}\}, \quad i = 1, \dots, N_f.$$

Для того чтобы найти оценки речевой передаточной функции [9, 10] между двумя дикторами для каждой частотной полосы, необходимо провести указанную процедуру для каждого диктора и вычислить матрицы \bar{X}_1 и \bar{X}_2 , а также векторы a_1 и a_2 . Тогда формула для вычисления речевой передаточной функции примет вид

$$|W_{12}(i)| = \frac{a_2(i)}{a_1(i)}, \quad i = 1, \dots, N_f.$$

В качестве эталона, относительно которого оценивается изменение речевых свойств, использован первый диктор. Аналогично можно опре-

делить речевую передаточную функцию между различными состояниями одного и того же диктора. Перевод полученных значений в децибелы выполняется по формуле [12]:

$$|W_{12}(i)|_{\text{дБ}} = 20 \lg |W_{12}(i)| = 20 \lg \frac{a_2(i)}{a_1(i)}, \quad i = 1, \dots, N_f.$$

Описание используемых речевых данных и предлагаемой методики идентификации человека-оператора по голосу. Для разработки методики идентификации человека-оператора по голосу использованы следующие речевые базы, созданные авторами статьи.

1. Речевая база, содержащая три различных слова (пилотаж, масштаб, навигация). Дикторы: девять летчиков вертолетной авиации с диагностированными заболеваниями органов слуха (номера 1–9) и 13 научных сотрудников ФГУП «ГосНИИАС» (номера 10–22). Всего — около 50 реализаций каждого слова для каждого диктора.

2. Речевая база, включающая в себя 20 различных слов (пилотаж, масштаб, навигация, тысяча, меньше, два, двадцать, взлет, пятьсот, ноль, двести, сто, десять, пять, пятьдесят, посадка, больше, руление, один, меньше). Дикторы: девять летчиков вертолетной авиации с диагностированными заболеваниями органов слуха (номера 1–9) и 10 научных сотрудников ФГУП «ГосНИИАС» (номера М1–М10). Всего — 30 реализаций каждого слова для каждого диктора.

Для возможности проведения сравнительного анализа из данных речевых баз использованы только реализации слов пилотаж, масштаб и навигация. Реализации слов произносились дикторами изолированно с паузами между словами 2...5 с, чтобы исключить влияние контекста. Первичная обработка заключалась в выделении каждой реализации слова в отдельный файл в звуковом формате WAV. Большинство известных средств автоматического разделения полезного речевого сигнала и внешних помех не отличается высокой точностью. Поэтому выделение полезного сигнала проводилось вручную средствами универсальной программы работы с аудиоданными *Audacity*®. При этом полезный речевой сигнал выделялся вместе с полями не более 0,1 с. Следует отметить, что визуальное сопоставление предварительно нарезанных вручную реализаций слов из обеих баз речевых данных показало: качество нарезки в первой базе оказалось заметно выше. Так, некоторые реализации из речевой базы на 20 слов имеют чрезмерно широкие поля или, наоборот, слишком сильно обрезаны, что привело к потере некоторой части полезного речевого материала.

Для предварительной обработки речевых сигналов использован упомянутый выше спектральный метод параметризации сигнала [11]. К полученным параметрическим портретам реализаций изолированных слов применен алгоритм вычисления речевой передаточной функции диктора [9, 10], причем в качестве эталонного использованы речевые данные диктора, которые не входили в упомянутые ранее речевые базы. Отметим, что передаточная функция вычислялась для отдельных реализаций слов, чтобы обеспечить возможность дальнейшей обработки каждой реализации обособленно от остальных. Таким образом, входными данными для алгоритмов классификации [13–15] являлись речевые передаточные функции для отдельных реализаций слов, представляющие собой векторы размерностью 36 (в соответствии с разделением речевого сигнала на 36 частотных полос).

Часть входных речевых данных рассмотрена в качестве обучающей выборки, для которой было априорно известно, какому диктору принадлежит каждая передаточная функция. По оставшимся данным проводилась идентификация личности диктора.

Экспериментальные результаты исследования возможности идентификации человека-оператора по его речевым характеристикам. Как правило, экипаж летательного аппарата состоит из командира воздушного судна, второго пилота и бортинженера, поэтому решать задачу идентификации личности целесообразно для небольшой группы. Принятая в работе минимальная численность исследуемой группы — четыре человека.

Результаты идентификации оператора для групп пилотов и научных сотрудников при классификации по одному слову приведены в табл. 1. Классификация и предварительное обучение модели проведены с помощью метода k ближайших соседей только по одному из трех слов, имеющих в речевой базе — пилотаж («п»), масштаб («м») или навигация («н») (см. табл. 1). При этом рассмотрены случаи разной размерности обучающей выборки (15 и 30 реализаций одного слова). В строках таблиц указано число и процент ошибок идентификации для каждого оператора и для всех операторов вместе при решении задачи по каждому из трех слов.

Анализ полученных результатов показывает, что использование большего числа реализаций слов при обучении модели приводит к уменьшению процента ошибок идентификации как минимум в 2,5–3 раза. Следует отметить более высокое качество идентификации личности в группах летчиков с диагнозом тугоухость по сравнению с группами научных сотрудников. Это показывает, что летчики обладают большей степенью разнообразия ха-

рактических речей [16, 17]. Высокий процент ошибок идентификации, полученных при работе с речевой базой, содержащей 20 слов, объясняется, как уже отмечено ранее, более низким качеством детектирования полезного речевого сигнала для реализаций из этой выборки. Процент ошибок идентификации летчиков с диагнозом тугоухость на речевой базе из трех слов (хорошая нарезка) при объеме обучающей выборки 30 реализаций не превысила 1 % при решении задачи в случае классификации по одному заранее заданному слову. Для научных сотрудников аналогичный процент ошибок составляет 1,0...2,5 % и может быть уменьшен в результате увеличения размерности обучающей выборки.

Таблица 1

**Результаты идентификации оператора для групп пилотов
и научных сотрудников при классификации по одному слову
(в числителе указано число ошибок идентификации,
в знаменателе — процент ошибок идентификации)**

Обучение по слову	Номер оператора				Всего
	11	12	13	14	
<i>Речевая база — три слова, научные сотрудники. Обучение модели по 15 реализациям</i>					
«П»	0/0	0/0	1/2,6	0/0	1/0,6
«М»	1/3,4	0/0	3/10,3	1/3,4	5/4,3
«Н»	1/2,6	0/0	0/0	4/10,5	5/3,2
<i>Речевая база — три слова, научные сотрудники. Обучение модели по 30 реализациям</i>					
«П»	0/0	0/0	0/0	0/0	0/0
«М»	0/0	0/0	1/7,1	0/0	1/1,7
«Н»	0/0	0/0	0/0	0/0	0/0
<i>Речевая база — три слова, летчики с диагнозом тугоухость. Обучение модели по 30 реализациям</i>					
Обучение по слову	1	2	3	4	Всего
«П»	0/0	0/0	0/0	0/0	0/0
«М»	0/0	0/0	0/0	0/0	0/0
«Н»	0/0	0/0	0/0	0/0	0/0
<i>Речевая база — 20 слов, летчики с диагнозом тугоухость. Обучение модели по 15 реализациям</i>					
Обучение по слову	2	3	4	5	Всего
«П»	1/6,6	1/6,6	2/13,3	0/0	4/6,6
«М»	1/6,6	0/0	2/13,3	0/0	3/5
«Н»	2/13,3	0/0	4/26,6	0/0	6/10

Окончание табл. 1

Обучение по слову	Номер оператора				Всего
	M1	M2	M3	M4	
<i>Речевая база — 20 слов, научные сотрудники. Обучение модели по 15 реализациям</i>					
«П»	0/0	1/6,6	0/0	2/13,3	3/5
«М»	0/0	0/0	0/0	0/0	0/0
«Н»	1/6,6	0/0	0/0	4/26,6	5/8,3

Результаты идентификации человека-оператора для групп пилотов и научных сотрудников при классификации по двум или трем словам приведены в табл. 2. Классификация и предварительное обучение модели проведены с помощью метода k ближайших соседей по двум или трем словам, имеющимся в речевой базе — пилотаж («п»), масштаб («м») или навигация («н»). Например, «п» + «м» обозначает, что обучение модели проводилось для слов «пилотаж» и «масштаб». В строках таблиц приведено число и процент ошибок идентификации для каждого оператора и для всех операторов вместе при решении задачи для каждого варианта.

Таблица 2

Результаты идентификации человека-оператора для групп пилотов и научных сотрудников при классификации по двум или трем словам (в числителе указано число ошибок идентификации для каждого оператора и для всех операторов вместе при решении задачи для каждого варианта, в знаменателе — процент ошибок идентификации)

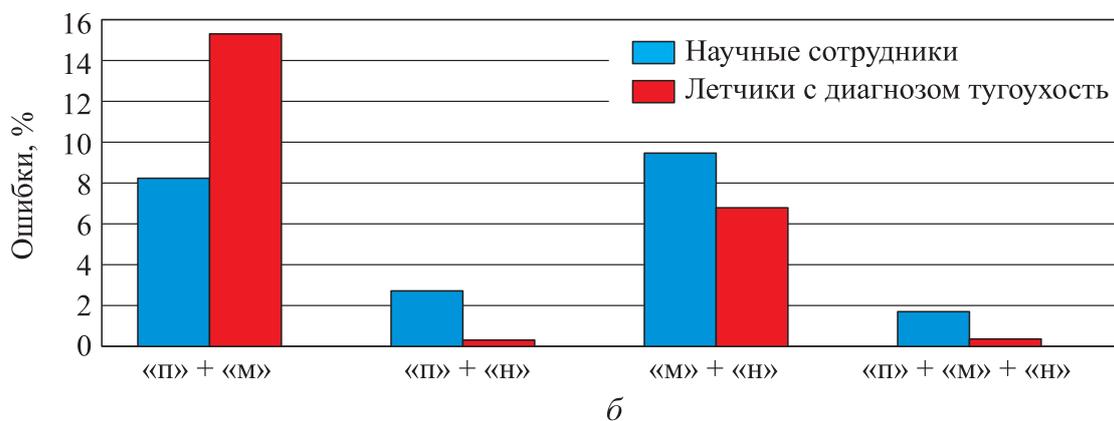
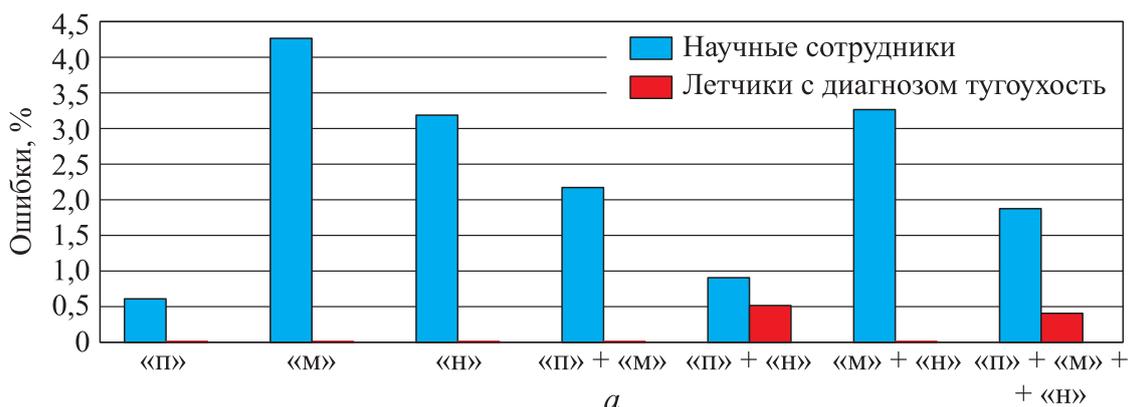
Обучение по слову	Номер оператора				Всего
	2	3	4	5	
<i>Речевая база — три слова, летчики с диагнозом тугоухость. Классификация по тем же словам, по которым проходит обучение</i>					
«П» + «М»	0/0	0/0	0/0	0/0	0/0
«П» + «Н»	0/0	0/0	0/0	1/2,1	1/0,5
«М» + «Н»	0/0	0/0	0/0	0/0	0/0
«П» + «М» + «Н»	0/0	0/0	0/0	1/1,6	1/0,4
<i>Речевая база — 20 слов, летчики с диагнозом тугоухость. Классификация по тем же словам, по которым проходит обучение</i>					
«П» + «М»	4/13,3	1/3,3	4/13,3	0/0	9/7,5
«П» + «Н»	3/10	1/3,3	6/20	0/0	10/8,3
«М» + «Н»	5/16,6	0/0	6/20	0/0	11/9,1
«П» + «М» + «Н»	6/13,3	1/2,2	8/17,7	0/0	15/8,3

Продолжение табл. 2

Обучение по слову	Номер оператора				Всего
	2	3	4	5	
<i>Речевая база — три слова, летчики с диагнозом тугоухость. Классификация по трем словам</i>					
«П» + «М»	0/0	15/25	23/38,3	0/0	38/15,8
«П» + «Н»	0/0	0/0	0/0	1/1,6	1/0,4
«М» + «Н»	2/3,3	0/0	1/1,6	14/23,3	17/7
«П» + «М» + «Н»	0/0	0/0	0/0	1/1,6	1/0,4
<i>Речевая база — 20 слов, летчики с диагнозом тугоухость. Классификация по трем словам</i>					
«П» + «М»	5/11,1	10/22,2	15/33,3	1/2,2	31/17,2
«П» + «Н»	15/33,3	8/17,7	9/20	2/4,4	34/18,8
«М» + «Н»	9/20	1/2,2	7/15,5	0/0	17/9,4
«П» + «М» + «Н»	6/13,3	1/2,2	8/17,7	0/0	15/8,3
<i>Речевая база — три слова, научные сотрудники. Классификация по тем же словам, по которым проходит обучение</i>					
Обучение по слову	11	12	12	14	Всего
«П» + «М»	3/4,4	0/0	3/4,4	0/0	6/2,2
«П» + «Н»	2/2,6	0/0	1/1,3	0/0	3/0,9
«М» + «Н»	3/4,4	0/0	2/2,9	4/5,9	9/3,3
«П» + «М» + «Н»	4/3,8	1/0,9	3/2,8	0/0	8/1,9
<i>Речевая база — 20 слов, научные сотрудники. Классификация по тем же словам, по которым проходит обучение</i>					
Обучение по слову	М1	М2	М3	М4	Всего
«П» + «М»	0/0	0/0	0/0	0/0	0/0
«П» + «Н»	1/3,3	0/0	0/0	0/0	1/0,8
«М» + «Н»	1/3,3	4/13,3	1/3,3	0/0	6/5
«П» + «М» + «Н»	2/4,4	1/2,2	1/2,2	0/0	4/2,2
<i>Речевая база — три слова, научные сотрудники. Классификация по трем словам</i>					
Обучение по слову	11	12	12	14	Всего
«П» + «М»	4/3,8	2/1,9	30/28,5	0/0	36/8,5
«П» + «Н»	9/8,5	0/0	3/2,8	0/0	12/2,8
«М» + «Н»	28/26,6	0/0	2/1,9	11/10,4	41/9,7
«П» + «М» + «Н»	4/3,8	1/0,9	3/2,8	0/0	8/1,9

Обучение по слову	Номер оператора				Всего
	М1	М2	М3	М4	
<i>Речевая база — 20 слов, научные сотрудники.</i>					
<i>Классификация по трем словам</i>					
«П» + «М»	7/15,5	1/2,2	1/2,2	1/2,2	10/5,5
«П» + «Н»	3/6,6	4/8,8	0/0	0/0	7/3,8
«М» + «Н»	2/4,4	8/17,7	4/8,8	0/0	14/7,7
«П» + «М» + «Н»	2/4,4	1/2,2	1/2,2	0/0	4/2,2
Обучение модели проводилось по 15 реализациям.					

Сравнительные результаты качества идентификации для групп операторов с нормальным слухом и летчиков с диагнозом тугоухость при обучении модели по 15 реализациям на речевой базе из трех различных слов приведены на рисунке.



Сравнительные гистограммы качества идентификации для групп операторов с нормальным слухом и летчиков с диагнозом тугоухость по тем же словам, по которым проводилось обучение модели (а), и по трем словам (б)

Анализ полученных результатов показывает, что использование при классификации слов, реализации которых не входили в обучающую выборку, приводит к увеличению процента ошибок идентификации по меньшей мере в 3 раза. Однако когда классификация проводилась по реализациям тех же слов, которые участвовали в обучении модели, сохранялось высокое качество идентификации с процентом ошибок около 0,5...3,0 % в зависимости от выбора группы операторов. Как и для случая классификации по реализациям одного слова, результаты групп летчиков с диагнозом тугоухость демонстрируют уменьшение процента ошибок идентификации в несколько раз по сравнению с группами научных сотрудников. Процент ошибок идентификации летчиков с диагнозом тугоухость на речевой базе из трех слов (хорошая нарезка) при объеме обучающей выборки 15 реализаций не превышает 0,5 % при решении задачи в случае классификации по тем же заранее заданным словам, по которым выполнялось обучение. Для научных сотрудников аналогичный процент ошибок составляет 1,0...3,5 %. Использование в задаче классификации слов, которые отсутствовали в обучающей выборке, приводит к увеличению процента ошибок идентификации до 15 %.

Заключение. Предложена методика голосовой идентификации летчика в составе экипажа на основе понятия речевой передаточной функции диктора по частоте. Представлены результаты экспериментальных оценок качества идентификации на речевых базах данных различной размерности и для разных групп операторов. Приведенные результаты показывают возможность идентификации оператора по небольшому числу кодовых слов с точностью около 97...99 %. Сравнительный анализ результатов апробации предложенной методики для различных групп операторов показывает, что группа летчиков с диагнозом тугоухость характеризуется большим разнообразием речевых характеристик по сравнению с группой научных сотрудников без диагностированных заболеваний органов слуха.

ЛИТЕРАТУРА

- [1] Желтов С.Ю., Себряков Г.Г. Моделирование многофункциональной деятельности человека-оператора при управлении летательным аппаратом. *XXVII Всерос. науч.-техн. конф. школы-семинара «Передача, прием, обработка и отображение информации о быстропротекающих процессах»*. М., ИД Академии им. Н.Е. Жуковского, 2016, с. 266–275.
- [2] Rabiner L., Juang B.H. *Fundamentals of speech recognition*. Prentice Hall, 1993.
- [3] Аграновский А.В., Леднов Д.А., Репалов С.А. Метод текстонезависимой идентификации диктора на основе индивидуальности произношения гласных звуков. *Акустика и прикладная лингвистика*, 2002, № 3, с. 103–115.

- [4] Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. М., Радио и связь, 2004.
- [5] Зубова П.И., Коваль С.Л. Идентификация личности по голосу и звучащей речи на основе комплексного анализа фонограмм. *Теория и практика судебной экспертизы*, 2007, № 3 (7), с. 68–76.
- [6] Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор. *Информационные процессы*, 2012, т. 12, № 1. URL: <http://www.jip.ru/2012/1-30-2012.pdf>
- [7] Greenberg C.S., Martin A.F., Przybocki M.A. Human assisted speaker recognition. URL: https://www.nist.gov/sites/default/files/documents/oles/8-Greenberg_Craig-Human-Assisted-Speaker-Recognition.pdf (дата обращения: 15.02.2019).
- [8] Евстигнеев Д.А., Копысов В.Х. Психологические основы взаимодействия в экипаже. Ульяновск, УВАУ ГА, 2007.
- [9] Korsun O.N., Mikhaylov E.I., Nakhaev M.Z. Speech spectral transfer function. *ITM Web Conf.*, 2017, vol. 10. DOI: <https://doi.org/10.1051/itmconf/20171001006>
- [10] Корсун О.Н., Нахаев М.З. Речевая передаточная функция диктора по частоте. *ВКИТ*, 2016, № 10 (148), с. 11–16.
- [11] Корсун О.Н., Михайлов Е.И. Метод выделения речевого сигнала, основанный на технологии контрольных карт. *Вестник компьютерных и информационных технологий*, 2017, № 9 (159), с. 24–31.
- [12] Корсун О.Н., Нахаев М.З. Понятие речевой передаточной функции диктора и экспериментальная оценка ее свойств. *Труды ГосНИИАС. Сер. Вопросы авионики*, 2016, № 1 (25), с. 33–38.
- [13] Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. *Springer Series in Statistics*. New York, Springer, 2009. DOI: 10.1007/978-0-387-84858-7
- [14] Nigsch F., Bender A., van Buuren B., et al. Melting point prediction employing k -nearest neighbor algorithms and genetic parameter optimization. *J. Chem. Inf. Model.*, 2006, vol. 46, no. 6, pp. 2412–2422. DOI: <https://doi.org/10.1021/ci060149f>
- [15] Воронцов К.В. Алгоритмы кластеризации и многомерного шкалирования. М., МГУ, 2007.
- [16] Корсун О.Н., Михайлов Е.И., Нахаев М.З. Результаты исследований параметров речи летчиков вертолетной авиации. *Научные чтения по авиации, посвященные памяти Н.Е. Жуковского*, 2014, № 2, с. 158–161.
- [17] Корсун О.Н., Иванов А.И., Михайлов Е.И. Исследование особенностей речи пилотов с диагнозом «тугоухость». *XXIV Всерос. науч.-техн. конф. школы-семинара «Передача, прием, обработка и отображение информации о быстропротекающих процессах»*. М., РПА АПР, 2013, с. 428–438.

Корсун Олег Николаевич — д-р техн. наук, начальник лаборатории ФГУП «ГосНИИАС» (Российская Федерация, 125319, Москва, ул. Викторенко, д. 7).

Михайлов Евгений Игоревич — аспирант МФТИ (Российская Федерация, 141701, Московская обл., Долгопрудный, Институтский переулок, д. 9).

Просьба ссылаться на эту статью следующим образом:

Корсун О.Н., Михайлов Е.И. Идентификация летчика в составе экипажа на основе речевой передаточной функции. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2019, № 5, с. 35–48. DOI: 10.18698/0236-3933-2019-5-35-48

IDENTIFICATION OF THE PILOT AS PART OF THE CREW USING SPEECH SPECTRAL TRANSFER FUNCTION

O.N. Korsun¹

marmotto@rambler.ru

E.I. Mikhaylov²

mihevkonig@mail.ru

¹ FSUE State Research Institute of Aviation Systems, Moscow, Russian Federation

² Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow Region, Russian Federation

Abstract

The paper deals with the problem of voice identification of the pilot as part of the crew, which is one of the ways to improve the interface of the cockpit of a modern aircraft. The main trends of pilot voice identification in the task of improving the cockpit interface are voice control of onboard equipment and accident investigation. We introduce a method for identifying the speaker's personality using the speaker's voice transfer function by frequency and the k -nearest neighbors data classification algorithm. Due to the nature of the task, identification was carried out for small groups of operators of up to four people. The main results of testing the proposed method on the experimental speech data that include 3 and 20 different isolated words are given. Findings of research show that the operator can be identified by a small number of code words with an accuracy of about 97–99 % when using the speaker's voice transfer function by frequency. The paper also presents a comparison of the results of applying the methodology for pilots of helicopter aviation with the diagnosis of hearing loss and for a group of operators without diseases of the organs of hearing

Keywords

Human-operator identification, speech spectral transfer function, frequency, classification, k-nearest neighbors method

Received 10.04.2019

© Author(s), 2019

This work was supported by the Russian Foundation for Basic Research (RFBR project no. 18-08-01142-a)

REFERENCES

- [1] Zheltov S.Yu., Sebryakov G.G. [Simulation of multi-functional human-operator activity while driving aircraft]. *XXVII Vseros. nauch.-tekh. konf. shkoly-seminara "Peredacha, priem, obrabotka i otobrazhenie informatsii o bystroprotekayushchikh protsessakh"* [XXVII Russ. Sci.-Tech. Conf. of School-Seminar "Transmission, receiving, processing and displaying information on high-speed processes"]. Moscow, ID Akademii im. N.E. Zhukovskogo Publ., 2016, pp. 266–275 (in Russ.).
- [2] Rabiner L., Juang B.H. Fundamentals of speech recognition. Prentice Hall, 1993.
- [3] Agranovskiy A.B., Lednov D.A., Repalov S.A. Method of text-independent speaker identification based on individuality of vowels pronouncing. *Akustika i prikladnaya lingvistika*, 2002, no. 3, pp. 103–115 (in Russ.).
- [4] Agranovskiy A.V., Lednov D.A. Teoreticheskie aspekty algoritmov obrabotki i klassifikatsii rechevykh signalov [Theoretical aspects of processing and classification algorithms for speech signals]. Moscow, Radio i svyaz Publ., 2004.
- [5] Zubova P.I., Koval S.L. Person identification by voice and speech based on complex analysis of phonograms. *Teoriya i praktika sudebnoy ekspertizy* [Theory and Practice of Forensic Science], 2007, no. 3 (7), pp. 68–76 (in Russ.).
- [6] Sorokin V.N., Vyugin V.V., Tananykin A.A. Person recognition by voice: analytical review. *Informatsionnye protsessy* [Information Processes], 2012, vol. 12, no. 1 (in Russ.). Available at: <http://www.jip.ru/2012/1-30-2012.pdf>
- [7] Greenberg C.S., Martin A.F., Przybocki M.A. Human assisted speaker recognition. Available at: https://www.nist.gov/sites/default/files/documents/oles/8-Greenberg_Craig-Human-Assisted-Speaker-Recognition.pdf (accessed: 15.02.2019).
- [8] Evstigneev D.A., Kopysov V.Kh. Psikhologicheskie osnovy vzaimodeystviya v ekipazhe [Psychological basis of interaction in the crew]. Ulyanovsk, UVAU GA Publ., 2007.
- [9] Korsun O.N., Mikhaylov E.I., Nakhaev M.Z. Speech spectral transfer function. *ITM Web Conf.*, 2017, vol. 10. DOI: <https://doi.org/10.1051/itmconf/20171001006>
- [10] Korsun O.N., Nakhaev M.Z. The dictor speech transfer function in frequency domain. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies], 2016, no. 10 (148), pp. 11–16 (in Russ.).
- [11] Korsun O.N., Mikhaylov E.I. The method for speech detection based on control charts techniques. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies], 2017, no. 9 (159), pp. 24–31 (in Russ.).
- [12] Korsun O.N., Nakhaev M.Z. Concept of speaker speech transfer function and experimental assessment of its properties. *Trudy GosNIIAS. Ser. Voprosy avioniki*, 2016, no. 1 (25), pp. 33–38 (in Russ.).
- [13] Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. *Springer Series in Statistics*. New York, Springer, 2009. DOI: 10.1007/978-0-387-84858-7

[14] Nigsch F., Bender A., van Buuren B., et al. Melting point prediction employing k -nearest neighbor algorithms and genetic parameter optimization. *J. Chem. Inf. Model.*, 2006, vol. 46, no. 6, pp. 2412–2422. DOI: <https://doi.org/10.1021/ci060149f>

[15] Vorontsov K.V. Algoritmy klasterizatsii i mnogomernogo shkalirovaniya [Clustering and multidimensional scaling algorithms]. Moscow, MSU Publ., 2007.

[16] Korsun O.N., Mikhaylov E.I., Nakhaev M.Z. Rezultaty issledovaniy parametrov rechi letchikov vertoletnoy aviatsii [Results of studies on speech parameters of helicopter pilots]. *Nauchnye chteniya po aviatsii, posvyashchennye pamyati N.E. Zhukovskogo* [Scientific readings on aviation dedicated to memory of N.E. Zhukovsky], 2014, no. 2, pp. 158–161 (in Russ.).

[17] Korsun O.N., Ivanov A.I., Mikhaylov E.I. [Study on pilots speech features with “hardness of hearing” diacrisis]. *XXIV Vseros. nauch.-tekh. konf. shkoly-seminara “Peredacha, priem, obrabotka i otobrazhenie informatsii o bystroprotokayushchikh protsessakh”* [XXIV Russ. Sci.-Tech. Conf. of School-Seminar “Transmission, receiving, processing and displaying information on high-speed processes”]. Moscow, RPA APR Publ., 2013, pp. 428–438 (in Russ.).

Korsun O.N. — Dr. Sc. (Eng.), Head of Laboratory, FSUE State Research Institute of Aviation Systems (Viktorenko ul. 7, Moscow, 125319 Russian Federation).

Mikhaylov E.I. — Post-Graduate Student, Moscow Institute of Physics and Technology (State University) (Institutskiy pereulok 9, Dolgoprudny, Moscow Region, 141701 Russian Federation).

Please cite this article in English as:

Korsun O.N., Mikhaylov E.I. Identification of the pilot as part of the crew using speech spectral transfer function. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2019, no. 5, pp. 35–48 (in Russ.).

DOI: 10.18698/0236-3933-2019-5-35-48