СИСТЕМЫ РАСПОЗНАВАНИЯ ЛИЧНОСТЕЙ ПО ГОЛОСУ

УДК 004.93+57.087.1

Ю. Н. Матвеев

ИССЛЕДОВАНИЕ ИНФОРМАТИВНОСТИ ПРИЗНАКОВ РЕЧИ ДЛЯ СИСТЕМ АВТОМАТИЧЕСКОЙ ИДЕНТИФИКАЦИИ ДИКТОРОВ

Исследуется информативность речевых признаков наиболее популярных при создании автоматических систем идентификации дикторов. Эксперименты проводились на речевой базе данных, собранной в различных акустических условиях (широком диапазоне отношений сигнал/шум и уровней реверберации) и с использованием различных каналов записи.

Ключевые слова: *признаки речи, идентификация дикторов.*

Введение. Речевой сигнал существенно отличается от других акустических сигналов, так как произносится человеком для человека и служит для обмена информацией между людьми. Поэтому в системах распознавания личностей по голосу (распознавания дикторов) целью первичной обработки речевого сигнала является выделение признаков речи, специфичных для отдельных дикторов.

Наиболее распространенными речевыми признаками для систем идентификации дикторов являются [1]:

— частота основного тона;

— частота формант;

— кепстральные коэффициенты.

Первые два признака используются в основном в экспертных и полуавтоматических системах идентификации дикторов. В большинстве автоматических систем идентификации дикторов в качестве признаков используются векторы кепстральных коэффициентов:

— линейно-частотных кепстральных коэффициентов (LFCC, Linear-Frequency Cepstral Coefficients) или мэл-частотных кепстральных коэффициентов (MFCC, Mel-Frequency Cepstral Coefficients), получаемых по спектру Фурье [2];

— коэффициентов линейного предсказания (LPCC, Linear Prediction Cepstral Coefficients) [3];

— коэффициентов перцептивного линейного предсказания (PLP, Perceptual Linear Prediction) [4].

Наилучшим из критериев эффективности признаков является критерий разделимости классов, который связан с вероятностями ошибок классификатора. Поэтому для оценки информативности признаков или компактности пространств признаков распознаваемых голосов дикторов будет использоваться вероятностный критерий, связанный с величиной равновероятной ошибки (EER, Equal Error Rate), т.е. точкой равенства ошибок первого и второго рода, определяемой по пересечению кривых распределений вероятностей этих ошибок.

Значение EER характеризует в данном случае информативность признаков для текстонезависимой автоматической системы идентификации личности по речевому сигналу. Чем меньше значение EER, тем меньше перекрытие между кривыми ошибок первого и второго рода и тем компактнее пространства признаков.

Целью предлагаемой работы является оценка информативности различных кепстральных признаков для автоматической системы идентификации дикторов.

Оценка информативности речевых признаков на тестовой базе данных. С целью оценки информативности различных признаков для автоматической системы идентификации дикторов была использована речевая база данных [5], характеристики которой приведены в табл. 1.

Таблица 1

Параметр	Канал				
	1	2	3	4	5
Среднее значение ОСШ, дБ	35	20	40	8	4
Средний уровень реверберации, мс	250	300	200	650	1000
Количество фонограмм	377	548	398	352	817
Количество дикторов	76	123	72	80	105

В таблице 1 используются следующие обозначения каналов:

1) микрофонный канал (ближний микрофон — гарнитура), микрофон расположен на расстоянии не более 30 см от рта говорящего;

2) телефонный IP-канал;

3) телефонный GSM-канал;

4) микрофонный канал (удаленный микрофон), микрофон расположен на расстоянии 1—2 м от рта говорящего;

5) микрофонный канал (удаленный микрофон), микрофон расположен на расстоянии 2—4 м от рта говорящего.

Оценка информативности признаков проводилась с помощью автоматической системы идентификации дикторов, представленной на конкурс по распознаванию дикторов [6] NIST Speaker Recognition Evaluation (SRE) 2010, проведенный Институтом стандартов и технологий США (NIST).

В качестве исследуемых признаков были выбраны:

1) супервектор, составленный из 13 коэффициентов вектора MFCC, их 13 первых производных и их 13 вторых производных;

2) супервектор, составленный из 18 коэффициентов вектора LPCC и их 18 первых производных;

3) супервектор, составленный из 13 коэффициентов вектора PLP, их 13 первых производных и их 13 вторых производных.

В табл. 2 приведены результаты оценки информативности признаков на тестовой базе. Курсивом выделены минимальные значения EER, полужирным шрифтом — максимальные. Чем меньше значение EER, тем выше информативность признака.

Таблица 2

Признак	Канал				
	1	2	3	4	5
MFCC	4,0	*5,5*	**5,0**	10,0	21,5
LPCC	*3,0*	**8,5**	4,5	*6,0*	**26,5**
PLP	**5,0**	*5,5*	*3,5*	**12,0**	*17,5*

Анализ коррелированности признаков речевых признаков. Опыт участия в конкурсе NIST SRE-2010 [6] показал, что большинство мировых лидеров в своих системах используют не отдельные признаки, а их комбинации. При этом наблюдалось повышение эффек-

тивности идентификации даже при наличии корреляции между смешиваемыми признаками. Таким образом, при совместном использовании различных наборов признаков дополнительным критерием информативности признаков может быть степень их некоррелированности с другими признаками набора.

Так, в работе [7] отмечается коррелированность различных кепстральных признаков. Исследовались производные этих признаков (дельта-характеристики) для учета временных изменений. Включение производных в вектор признаков позволяет снизить влияние мультипликативных искажений сигнала, в силу того что эти искажения обычно медленно изменяются во времени и аддитивны в кепстральной области.

Из табл. 3 следует, что LPCC-коэффициенты имеют сильную корреляцию с MFCC-коэффициентами. Как отмечается в работе [7], это ожидаемый результат, поскольку оба этих признака описывают огибающую спектра. Кроме того, производные параметры кепстра также имеют высокую корреляцию, что объясняется схожестью методов их вычисления: ΔLPCC есть производная LPCC.

Таблица 3

Корреляция наборов признаков

Признак	ΔMFCC	LPCC	ΔLPCC
MFCC	0,77	0,88	0,71
ΔMFCC	—	0,73	0,69
LPCC	—	—	0,85

В работе [7] приведены результаты экспериментов по сравнению ряда других признаков, в том числе MFCC и PLP. Эксперименты проводились с использованием классификатора на основе смесей гауссовых распределений различного порядка (в зависимости от объема обучающих данных). Результаты исследований, приведенные в табл. 4, показали, что PLP не имеет преимуществ перед MFCC.

Таблица 4

Надежность идентификации
(в процентах правильно идентифицированных дикторов)

Порядок модели	MFCC	PLP
2	95,36	82,26
4	97,14	93,93
8	98,33	96,79
16	99,52	98,10
32	99,05	98,45

В обзоре [8] сделан вывод о том, что различные кепстральные признаки, такие как MFCC, LFCC, LPCC и PLP, имеют сильную корреляцию. Однако возможно их комбинирование (смешивание) для повышения надежности идентификации [7].

В табл. 5 дана оценка средней корреляции (СКО = 0,01) признаков по каналам 1—4 тестовой базы данных (см. табл. 1). Наиболее коррелированными признаками снова оказались MFCC и LPCC, а наименее — LPCC и PLP. Полученное значение корреляции признаков MFCC и LPCC согласуется с полученным в работе [7] и приведенным в табл. 3.

Таблица 5

Признак	LPCC	PLP
MFCC	0,84	0,81
LPCC		0,69

В табл. 6 дана оценка средней корреляции (СКО = 0,01) признаков по каналу 5 тестовой базы данных (см. табл. 1). Данный канал характеризуется высоким уровнем реверберации (более 1000 мс) и низким соотношением сигнал-шум (4 дБ). В таких акустических условиях наиболее коррелированными оказались признаки MFCC и PLP, а наименее — LPCC и PLP.

Таблица 6

Признак	LPCC	PLP
MFCC	0,70	0,82
LPCC	—	0,57

В табл. 7 приведены результаты экспериментов по комбинированию признаков, которые согласуются с приведенными выше оценками.

Таблица 7

Признак	Канал									
	1		2		3		4		5	
	Вес	EER, %	Вес	EER, %	Вес	EER, %	Вес	EER, %	Вес	EER, %
MFCC	0,004	4,0	0,465	**5,5**	0,034	10,0	0,240	5,0	0,174	21,5
LPCC	0,790	**3,0**	0,005	8,5	0,766	**6,0**	0,220	4,5	0,136	26,0
PLP	0,206	5,0	0,530	**5,5**	0,200	12,0	0,542	**3,5**	0,690	**17,5**
Комбинация (смесь)	1	2,5	1	4,5	1	6,0	1	3,0	1	17,0

Из полученных результатов можно сделать следующие выводы:

1) комбинирование (смешивание) признаков всегда обеспечивает наименьшее значение EER;

2) признак, имеющий наименьшее значение EER, всегда имеет наибольший весовой коэффициент (вес);

3) признак PLP менее коррелирован с MFCC и LPCC, чем MFCC и LPCC между собой, поэтому он всегда имеет значимый вес;

4) признаки MFCC и LPCC имеют высокую степень корреляции, поэтому один из них часто вносит очень мало дополнительной информации в обобщенное решение.

Заключение. В настоящей работе исследована информативность широко известных наборов речевых признаков, таких как MFCC, LFCC, LPCC и PLP. В качестве критерия информативности для отбора признаков в системе идентификации дикторов по голосу использовалось значение EER.

Показано, что MFCC, LPCC и PLP имеют сильную корреляцию, а также, что ни один из рассмотренных признаков не дает преимуществ по сравнению с другими по уровню информативности в различных акустических условиях и в различных каналах записи. Однако возможно их комбинирование для повышения надежности идентификации дикторов по голосу. Результат смешивания признаков всегда обеспечивает наименьшее значение EER.

СПИСОК ЛИТЕРАТУРЫ

1. *Матвеев Ю. Н.* Технологии биометрической идентификации личности по голосу и другим модальностям // Вестн. МГТУ им. Н. Э. Баумана. Сер. Приборостроение. Специальный выпуск. Биометрические технологии. 2012. № 3(3). С. 46—61.

2. *Huang X., Acero A., Hon H.* Spoken Language Processing: A guide to theory, algorithm, and system development. Prentice Hall, 2001. 1008 p.

3. *Zheng F., Zhang G., Song Z.* Comparison of Different Implementations of MFCC // J. Computer Sci. and Techn. 2001. Vol. 16, N 6. P. 582—589.

4. *Hermansky H., Malayath N.* Speaker Verification Using Speaker-Specific Mappings // Proc. of the Workshop on Speaker Recognition and its Commercial and Forensic Applications. Avignon, 1998. P. 111—114.

5. База данных для идентификации говорящего по голосу “RUASTEN”. Регистрационное свидетельство № 2010620533 от 20.09.2010.

6. *Матвеев Ю. Н., Симончик К. К.* Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // Тр. 20-й Междунар. конф. по компьютерной графике и зрению „ГрафиКон'2010". СПб: СПбГУ ИТМО, 2010. С. 315—319.

7. *He W., Hong P.* The Application of Fusion Technology for Speaker Recognition // Intern. J. of Computer Science and Network Security. 2007. Vol. 7, N 12. P. 300—303.

8. *Kinnunen T., Li H.* An overview of text-independent speaker recognition: From features to supervectors // Speech Communication. 2010. Vol. 52, N 1. P. 12—40.

Сведения об авторе

Юрий Николаевич Матвеев — д-р техн. наук; ООО „ЦРТ-инновации", Санкт-Петербург; главный научный сотрудник; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра речевых информационных систем; профессор;
E-mail: matveev@mail.ifmo.ru

Рекомендована кафедрой
речевых информационных систем

Поступила в редакцию
22.10.12 г.