

Исследование признаков SDC (Shifted Delta Cepstra) в задаче идентификации языка

Д.В. Иванько, А.В. Козлов

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ООО «Центр Речевых Технологий»
ivanko@speechpro.com, kozlov-a@speechpro.com

Аннотация

В статье представлены результаты работы по изучению признаков SDC, используемых для целей идентификации языка. Несмотря на то, что первое упоминание о самих признаках SDC появилось довольно давно [3] отсутствие в первой работе, и в последующих, многих деталей препятствовало созданию более эффективной системы идентификации языка (СИЯ). Восстановление этих деталей, с экспериментальным подтверждением и представление полученных результатов является основной целью данной статьи.

1. Задача идентификации языка

Задача идентификации представляет собой довольно сложный комплекс подзадач физико-математического направления, результатом которого является система, способная с определённой долей вероятности отличить один язык от другого, а в случае более успешного решения, способна разделять и различные диалекты одного и того же языка.

В современном обществе, происходит непрерывное развитие и сравнение таких систем, стимулирующее изучение данной области на международном уровне Национальным институтом стандартов и технологий (The National Institute of Standards and Technology – NIST) [1-2] предоставляемым каждый раз оригинальные звуковые базы, для испытания созданных систем. Первые попытки распознавания языка проводились в 1996-ом году, стабильное сравнение СИЯ начало проводится с 2001 года, когда каждый научный центр или компания смогли посоревноваться на мировом уровне и представить свои достижения в области идентификации языка, раз в два года. В представляемых же результатах чаще всего отсутствует наименование участников, а только результаты идентификации языка, подчёркивая тем самым, что главной целью проводимых сравнений является не рекламирование, а лишь аде-

кватная оценка собственной позиции на мировом уровне и стимулирование развития данной области.

Ключевая же цель создания подобных СИЯ служит обеспечение более быстрой и эффективной коммуникации общества, информированности и готовности к взаимодействию.

2. Структура СИЯ SDC

Сами изучаемые признаки SDC, являются составной частью более сложной СИЯ. В данном разделе показывается само место используемых признаков в задаче идентификации языка, а также представляется найденный наиболее оптимальный вариант образования признаков SDC. Обоснование и экспериментальное подтверждение данного построения представляется в разделе 3.

2.1. Система используемая для изучения

Совокупность ключевых элементов, практически применяемых систем, можно разделить на три части [3]:

- используемые признаки;
- математический аппарат системы;
- критерий принятия решения.

Используемые признаки, это вектор тех информативных параметров сигнала, который был выделен из фонограммы, и по которому система будет идентифицировать языковую принадлежность.

Математическим аппаратом можно обозначить ту стадию обработки исходных параметров, благодаря которой совокупность фонограмм одного и того же языка оказывается объединённой между собой в какой либо части пространства и отделённой от фонограмм, принадлежащих другим языкам.

Критерий принятия решения, это то условие, по которому система разделяет и идентифицирует языковую принадлежность фонограммы.

В рамках проводимого исследования признаков SDC, производились внесения изменений только на первом уровне, с поиском наиболее оптимальной структуры образования вектора параметров.

По результату первых экспериментов было выяснено, что переобучение системы, перенастройка под новые параметры занимает довольно продолжительное время, что в свою очередь является ощути-

Интернет и современное общество: сборник научных статей XVI Всероссийской объединённой конференции IMS-2013, Санкт-Петербург, 9 - 11 октября 2013 г.

мым препятствием для свободных экспериментов с СИЯ.

Во избежание серьёзных временных затрат было решено собрать экспериментальную модель, отражающую все аспекты реальной существующей СИЯ, но требующую для своего переобучения существенно меньше времени. Произведено также обучение и ранее используемых признаков под экспериментальную модель, для более объективной оценки. Все последующие результаты экспериментов в разделе 3.2 были представлены для собранной экспериментальной модели.

2.2. Используемые признаки SDC

Вектор признаков SDC, используемый в системе идентификации языка, представляет собой, объединённый набор первых сглаженных производных MFCCs (Δ MFCCs) на довольно продолжительном временном интервале, к тому же впоследствии объединяемый с самими статичными признаками MFCCs.

В литературе [3-7], указывается, что данные признаки характеризуются четырьмя параметрами:

M-d-p-k. Признак M – представляет собой количество MFCCs коэффициентов, используемых для построения SDC-признаков, и впоследствии с которыми они конкатенируются, для получения результирующего вектора. Признак d – это степень сглаженной производной, p – это шаг, через который берутся вектора для построения признаков SDC. k – это количество, взятых векторов Δ MFCCs, для построения результирующего вектора.

На рисунке 1, представлена типичная схема построения признаков вектора SDC. В порядке сверху вниз представлена: осциллограмма фрагмента звукового сигнала, построенные по данной осциллограмме семь Мел-кепстральных коэффициентов и соответствующие им сглаженные производные первой степени.

За время использования признаков SDC наиболее оптимальным решением явилась комбинация 7-1-3-7 SDC, финальной размерностью 56. Для построения подобных признаков подходит предоставляемые скрипты MATLAB Central [3].

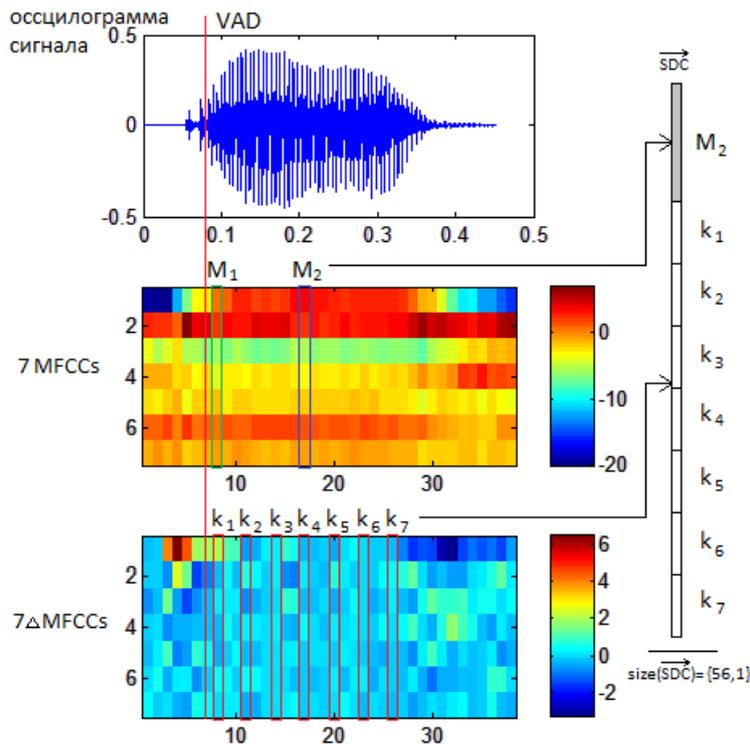


Рис. 1. Образование вектора признаков SDC

В результате представленных в разделе 3 экспериментов, была найдена на текущий момент наиболее удачная комбинация применения признаков SDC и их интеграции с VAD. Данная комбинация и представлена далее. После построения признаков MFCCs и Δ MFCCs, они конкатенируются в результирующий вектор SDC, следующим образом:

- по результату разметки VAD, выделяются те вектора построенного сигнала, которые относятся к речи;
- в результирующий вектор включается ближайший вектор Δ MFCCs, каждый из отсчётов которого создан на сигнале, который прошёл VAD;
- с ним конкатенируется следующий через p-векторов, вектор Δ MFCCs;

- таким образом, собирается k векторов;
- после этого, собранные вектора конкатенируются со статичными MFCCs, но взятыми из центра области, которая использовалась для построения Δ MFCCs.

Подобным образом образуется первый вектор речевого сигнала, по признакам SDC. Все после-

дующие вектора образуются аналогично. Вопрос об исключении векторов, после их выхода за пределы разметки VAD также рассматривался, и составляет отдельную группу экспериментов, представленную в разделе 3.1.

Таблица 1. Результаты экспериментов

№	Используемые признаки СИЯ	Результаты эксперимента	
		EER_Train, %	EER_Test,%
Сравнение по СИЯ			
1	MFCCs 39	-	8.24
2	7-1-3-7 SDC cdl	-	11.39
3	7-1-3-7 SDC cds	-	27.48
Сравнение по экспериментальной модели			
4	MFCCs 39	5,22	10.73
5	7-1-3-7 SDC cdl	7,03	13,00
6	7-1-3-7 SDC utb	11,27	19,76
7	7-1-3-7 SDC cdl + cM+0MFCCs+N	3,50	8,00
8	7-1-3-7 SDC cdl + cM+0MFCCs+uN	3,50	7,78
9	7-1-3-7 SDC cdl +0MFCCs+uN	4,17	9,00
10	7-1-3-7 SDC cMl + cM+0MFCCs+uN	3,36	7,44

3. Результаты экспериментов

По представленным в оригинальных статьях текстам по применению признаков SDC, было выделено несколько направлений для последующего исследования. По результатам проведенных экспериментов был выделен второй этап экспериментов, но уже по поиску наиболее удачной комбинации признаков. Исследования на втором этапе проводились с экспериментальной моделью.

3.1. Поиск наиболее оптимальной интеграции SDC и VAD

На данном этапе, была собрана рабочая СИЯ с признаками 7-1-3-7 SDC, в полном соответствии с системой, используемой на практике, а признаки соответствовали изученной оригинальной статье [3]. Основной вопрос неизвестный в данной системе являлся принцип интеграции SDC и VAD, поскольку признаки строятся на довольно продолжительном, для произнесения интервале времени. Для исследования данной области были собраны четыре различные возможные схемы использования SDC и VAD. Эти способы подразделялись на два класса с включением векторов Δ MFCCs которые выходили за разметку VAD (cdl) и без этого включения (cds), вторым классом являлось использование производных, которые построены по векторам находящимся за пределами VAD (cMl, вариант cMs не рассматривается т.к. первые же эксперименты показали не состоятельность подобного построения).

Отдельным способом комбинирования может считаться метод utb – первоначальное объединение

всех векторов прошедших VAD, однако как показали результаты эксперимента 6, подобная комбинация является ошибочной.

3.1.1. Ход улучшения SDC

После анализа полученных результатов (эксперименты под номерами 1-6, таблицы 1) стало очевидным, что представленная комбинация признаков нуждается в доработке, поскольку, несмотря на увеличенную размерность признаков и сложность системы на практике оказывается значительно хуже используемых ранее признаков MFCCs. В литературе [5], были даны кое-какие практические рекомендации, и намёки на использование нормализации признаков. Первые несколько не очень удачных экспериментов не приводятся, однако после серии подобных исследований была собрана наиболее удачная комбинация, принцип построения которой и указан в разделе 2. Комбинация под номером 7, обозначающая:

- центрирование MFCCs вектора (+cM);
- включение нулевого коэффициента, обычно не используемого при подсчёте MFCCs (+ 0 MFCCs);
- нормализация полученных признаков, с приведением их к $N(0,1)$, обозначаемая, при подсчёте по всему речевому содержанию (+N).

Центрирование MFCCs вектора представляет собой, при ближайшем рассмотрении на рисунке 1, конкатенацию в результирующем векторе, MFCCs-вектора находящегося в центре семи векторов SDC (вектор M_2 , на рисунке 1). В отличие от стандартного способа конкатенации, где признаки соединяются

начиная с того MFCCs-вектора, от которого они начинают своё образование (вектор M_1 на рисунке 1). Подобная перестановка нигде не упоминается, по крайней мере, во встречаемой литературе, однако экспериментальная состоятельность подобной перестановки подтверждается сравнением случаев 8 и 9 таблицы 1, где единственное различие представляет собой центрирование признака (+сМ).

Включение нулевого коэффициента упоминалось в работе [5], что оказало существенный вклад на результаты работы. Нормализация признаков также упоминалась во многих работах по SDC, однако в отличие от используемого CMS (Central Mean Subtraction), происходит не только вычитание среднего, но и приведение набора признаков к единичной дисперсии, т.е. приведение к $N(0,1)$.

В той же работе [5], было дано упоминание о так названном «utterance normalization», для испытания подобной нормализации по определённым высказываниям был проведён эксперимент 8 (+uN), который показал состоятельность подобного преобразования признаков. Сущность его заключалась в том что, при продолжении паузы в течении более T секунд (в эксперименте 8, $T=9$ с.) происходит разделение речевого сегмента, и на каждом из сегментов происходит отдельная нормализация к $N(0,1)$.

В заключении проводился эксперимент по расширению границ области включения векторов, используемых для обработки сигнала, описываемый в части 3.1, именуемый сМ1, который показал не только улучшение показателей, но и из чего можно сделать вывод о зависимости между качеством идентификации языка и включаемых в обработку векторов. Однако подобное улучшение будет иметь сильную связь с качеством, используемого VAD, поэтому необходимо будет это учитывать, при практическом использовании признаков SDC.

Литература

- [1] Козлов А.В., Кудашев О.Ю., Матвеев Ю.Н., Пеховский Т.С., Симончик К.К., Шулипа А.К., Система идентификации дикторов по голосу для конкурса NIST SRE 2012 // Труды СПИИРАН, 2013.
- [2] Матвеев Ю.Н., Симончик К.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // Труды 20-й международной Конференции по Компьютерной Графике и Зрению «ГрафиКон'2010» Санкт-Петербург, Россия СПбГУ ИТМО, 2010.
- [3] Torres-Carrasquillo P., Singer S., Kohler M., Richard J., Reynolds D., Deller J. Approaches to Language Identification using Gaussian Mixture Models and Shifted Delta Cepstral Features, 2002.
- [4] Schwarz P. Phoneme Recognition Based on Long Temporal Context, Brno 2008.
- [5] Rong T. Automatic Speaker and Language Identification // A First Year Report Submitted to the School of Computer Engineering of the Nanyang Technological University, 2006.

- [6] Shifted Delta Coefficients (SDC) Computation from Mel Frequency Cepstral Coefficients (MFCC) // MATLAB Central URL: <http://www.mathworks.com/matlabcentral/fileexchange/31478-shifted-delta-coefficients-sdc-computation-from-mel-frequency-cepstral-coefficients-mfcc> (дата обращения: 15.08.2013).

- [7] Hubeika V., Lukas B., Matejka B., Schwarz P. Discriminative Training and Channel Compensation for Acoustic Language Recognition.

The Research of Shifted Delta Cepstral (SDC) features problem of language identification

D.V. Ivanko, A.V. Kozlov

This article describes the results of research SDC features used for the purpose of language identification. Describes in detail the construction of a feature vector SDC with best experimental evaluation in this article. Displays the results of experimentation to find a better way of interaction between SDC and VAD. Some improvement are also described. The result of this implementation and this interaction are also presented with experimental confirm.