

П.В. Малинин, В.В. Поляков

Влияние искажений голосового сигнала на идентификацию диктора

P.V. Malinin, V.V. Polyakov

Influence of the Vocal Signal Distortion on Speaker's Identification

Предложен метод оценки влияния искажений голосовых сигналов, основанный на проекционных методах анализа многомерных данных. Показана возможность определения уровня искажений, при котором сохраняется разделение голосов дикторов. Полученные результаты свидетельствуют о перспективности предложенного метода для повышения надежности идентификации личности по голосу.

Ключевые слова: идентификация диктора, метод главных компонент, защита информации.

При идентификации личности по голосу, проводимой в реальных условиях, встречается ряд серьезных затруднений [1]. Во-первых, возможны искажения, связанные непосредственно с диктором и обусловленные особенностями его психофизического состояния, заболеванием и т.п. Во-вторых, на голосовой сигнал накладываются внешние механические шумы, которые могут существенно его искажать. В силу этих обстоятельств актуальной задачей является изучение влияния искажений исходного сигнала на возможность надежной идентификации диктора. Для исследования этого влияния в настоящей работе используется подход, изложенный в [2, 3] и заключающийся в применении проекционных методов анализа многомерных данных к голосовым сигналам. Именно, привлекается метод главных компонент, в котором коэффициенты кепстрального разложения голосового сигнала используются для построения графика счетов [4, 5]. Каждая точка на таком графике соответствует одному голосовому сигналу конкретного диктора, взаимное расположение точек позволяет сопоставить неизвестный сигнал с диктором и тем самым провести идентификацию.

1. Влияние искажений, обусловленных изменением частотного спектра исходного сигнала. Для количественной оценки влияния искажений голосового сигнала на графике счетов для первых главных компонент выделялись кластеры точек, каждый из которых соответствовал одному конкретному диктору. Далее программно вычислялся центр масс для каждого такого кластера и по мак-

The article offers a method for estimating the influence of voice signals distortions based on the projection methods of multivariate data analysis. The authors demonstrate the possibility to determine the level of distortion, when division of speakers voices remains. The results indicate the prospects of the proposed method for improving the reliability of identification by voice.

Key words: speaker identification, multivariate data analysis, information security.

симально удаленной от центра масс точке устанавливался максимальный радиус – размер этого кластера. Проводилась передискретизация спектра исходного сигнала [6]. С помощью быстрого преобразования Фурье в исходный частотный спектр добавлялось определенное количество нулевых отсчетов, затем осуществлялось обратное преобразование Фурье, создававшее уже искаженный сигнал. Данные искажения аналогичны повышению частоты основного тона исходного голосового сигнала и изменению набора последующих гармоник. В качестве величины, характеризующей искажение, выступал коэффициент нелинейных искажений K , равный отношению среднеквадратичной суммы спектральных компонентов выходного сигнала, отсутствующих в спектре входного сигнала, к среднеквадратичной сумме спектральных компонентов входного сигнала. Проведенным преобразованиям соответствовали значения K , изменявшиеся в интервале от $K = 0,3$ до $K = 0,4$. На графике счетов искаженные сигналы изображались совокупностью точек, удалявшихся от области кластера исходных сигналов по мере увеличения K .

На рисунке 1 для иллюстрации приведен участок спектра исходного сигнала, охватывающий интервал частот от $f = 650$ Гц до $f = 700$ Гц, для которого искажения оказались наиболее заметными (A – амплитуда звуковых колебаний, A_{max} – максимальная амплитуда). На рисунке 1б для сопоставления представлен частотный спектр искаженного сигнала ($K = 0,34$).

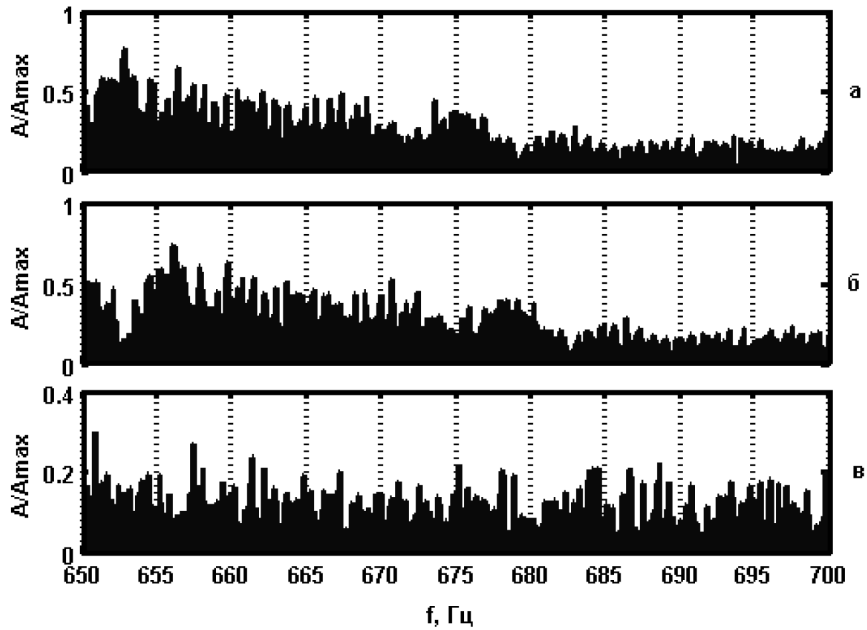


Рис. 1. Участок спектра звуковых сигналов: а – исходный голосовой сигнал; б – искаженный голосовой сигнал; в – использовавшийся для искажения сигнала шум

График счетов для одного типичного диктора (PC1 и PC2 – первая и вторая главные компоненты) изображен на рисунке 2. Окружность с центром в точке С, соответствующем центру масс, задает область кластера сигналов (точки 1) для данного диктора. Совокупность точек 2 описывает искаженные сигналы: точка 1 – $K = 0,32$, точка 2 – $K = 0,34$, точка 3 – $K = 0,36$, точка 4 – $K = 0,38$, точка 5 – $K = 0,40$.

Как видно из рисунка 2, при увеличении коэффициента нелинейных искажений наблюдается постепенное удаление точек, относящихся к искаженным сигналам, от кластера исходных точек. При значениях $K = 0,36$ эти звуковые сигналы уже не могут быть надежно идентифицированы как принадлежащие данному диктору.

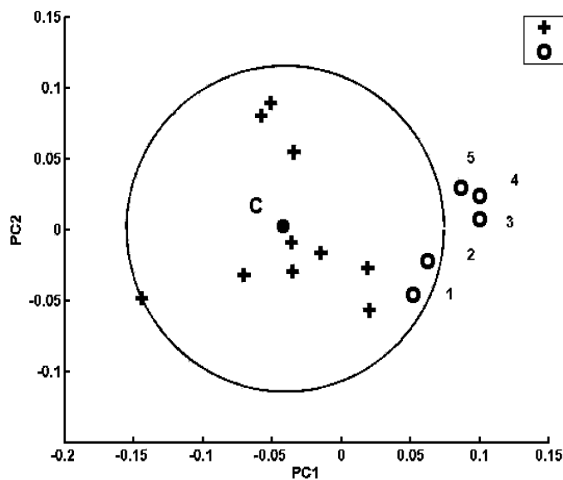


Рис. 2. Область счетов исходных и искаженных сигналов для одного диктора: 1 – исходные сигналы, 2 – искаженные сигналы (цифрами обозначены разные значения величины искажения)

2. Влияние искажений, обусловленных наложением случайных помех на исходный сигнал. Для изучения влияния случайных акустических помех программно задавался белый гауссовский шум [7], амплитуда которого выбиралась по соотношению «сигнал–шум». В качестве величины, характеризующей наложенный шум, выбирался относительный уровень шума L , равный отношению амплитуд шума и сигнала и изменявшийся в диапазоне от $L = 0,2$ до $L = 0,5$. Типичный участок спектра накладывавшихся помех приведен на рисунке 1в.

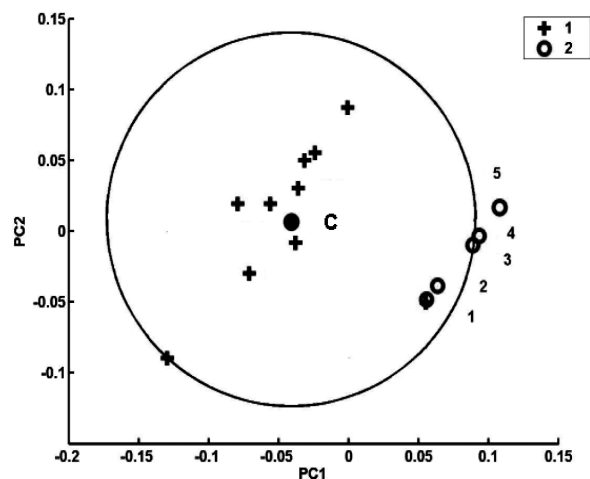


Рис. 3. Область счетов исходных и искаженных сигналов для одного диктора: 1 – исходные сигналы; 2 – сигналы с наложенным шумом (цифрами обозначены разные значения уровня шума)

На рисунке 3 изображен график счетов для одного диктора. Точки 1 соответствовали исходным неискаженным сигналам, точки 2 описывали совокупность сигналов с наложенными случайными помехами и соответствовали относительному уровню шума $L = 0,2$ (точка 1), $L = 0,25$ (точка 2), $L = 0,3$ (точка 3), $L = 0,4$ (точка 4), $L = 0,5$ (точка 5). Расположение точек 2 на рисунке 3 показывает, что при увеличении уровня шума наблюдается удаление точек, соответствовавших искаженным сигналам, из

области кластера неискаженных сигналов. При значениях $L \sim 0,5$ и выше надежная идентификация диктора представляется затруднительной.

Выводы. Приведенные в работе результаты позволяют выявить особенности влияния основных видов искажений голосового сигнала на идентификацию диктора. Учет этих особенностей должен обеспечить возможность повышения надежности такой идентификации.

Библиографический список

1. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу // Радио и связь. – М., 1981.
2. Малинин П.В., Поляков В.В. Иерархический подход в задаче идентификации личности по голосу с помощью проекционных методов классификации многомерных данных // Доклады ТУСУР. – 2010. – Ч. 1 (21).
3. Малинин П.В., Поляков В.В. Применение методов анализа многомерных данных в задаче идентификации личности по голосу // Известия АлтГУ. – 2010. – №1(65).
4. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: пер. с англ. / под ред. М.В. Назарова, Ю.Н. Прохорова. – М., 1981.
5. Эсбенсен К. Анализ многомерных данных. Избранные главы: пер. с англ. С.В. Кучерявского / под ред. О.Е. Родионовой / ИПХФ РАН. – Черногловка, 2005.
6. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов: пер. с англ. / под ред. С.Я. Шаца. – М., 1979.
7. Денда В. Шум как источник информации: пер. с нем. – М., 1993.