

БИОМЕТРИЯ: ПРЕДСКАЗАНИЕ ДЛИТЕЛЬНОСТИ ТОНАЛЬНЫХ ЗВУКОВ СЛИТНОЙ РЕЧИ

Калашиников Д.М.

В статье рассматривается проблема разделения на звуки слитной речи, определение их параметров и формирование баз звуков, характерных для голоса конкретного человека. Предложено строить распределение длин тональных звуков и по нему предсказывать наиболее вероятную длину одного звука слитной речи.

Ключевые слова: автокорреляционная функция, линейное предсказание речи, непокадровая обработка сигнала, смесь нормальных законов распределения, фрагментатор слитной речи

Введение

Механизм классификатора «тон/шум», построенный на линейном предсказании речи, на данный момент времени используется в большинстве современных вокодеров. К сожалению, качество принятия решений линейным предсказателем «тон/шум» низкое. Причина низкого качества состоит в покадровой обработке. В одном кадре может содержаться как тон, так и часть шума. Существует большая вероятность того, что автокорреляционная функция на таком промежутке даст на выходе неверное значение.

Классификация «тон/шум»

Автомат должен более надежно классифицировать фрагменты звука на тональные и шумовые. Для этого автомат должен обладать инфор-

мацией об ожидаемой длине звука в конкретной речи конкретного человека.

На рис. 1 приведен пример звукового файла с результатом классификации «тон/шум». Из рисунка видно, что появляются тональные участки разной длины. Алгоритм, осуществляющий данную классификацию, использует покадровую обработку сигнала, что вызывает случаи, на которых кадр может содержать одновременно тон и шум. На подобных участках кадра автокорреляционная функция становится неустойчивой, и велика ошибка в определении периода основного тона.

Чтобы свести эту ошибку к минимуму, необходимо четко определять границы тона, при этом каждый раз уточняя необходимые параметры. Было произведено исследование по работе алгоритма линейного предсказателя на классификацию «тон/шум» на нескольких фразах разных людей. При десяти повторных произношениях, представленной на рис. 1 фразы алгоритм выдал несколько отличных друг от друга участков «тон/шума».

В целом, все аудиозаписи выявили общие тоновые и шумовые участки, иногда подобные, иногда разделенные. В данном примере кадр был взят длиной в 256 отсчетов. Постоянная потеря участков такой длины искажает биометрию пользователя и сказывается в задачах аутентификации человека по голосу.

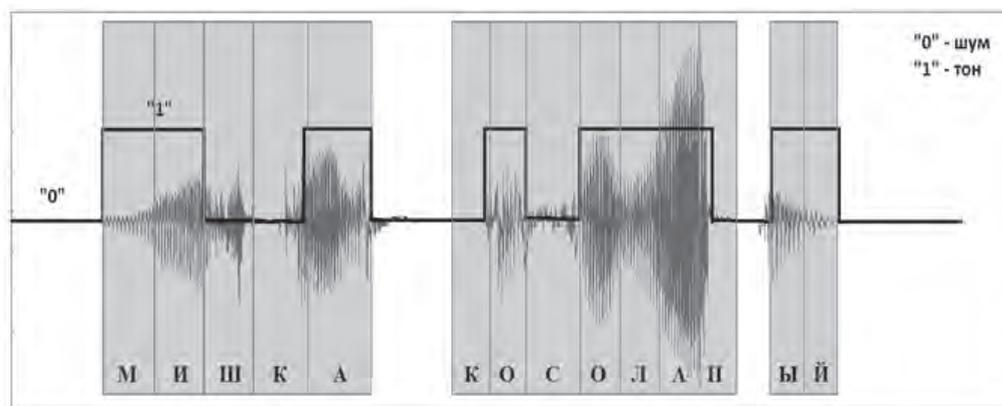


Рис. 1. Распределение «тон/шум» на фразе «Мишка косолапый»

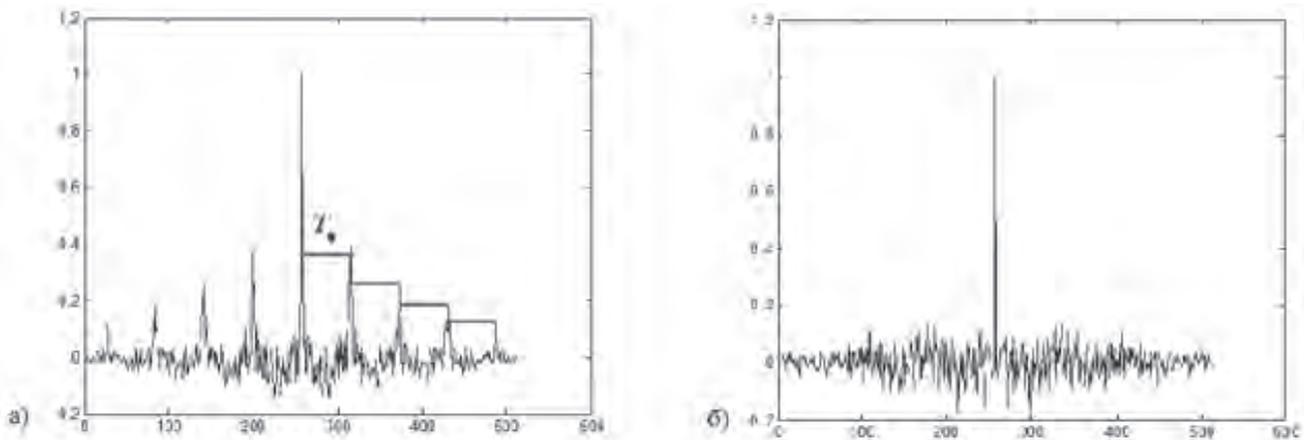


Рис. 2. Автокорреляционная функция на а) тоновом участке б) шумовом участке

На рис. 2 приведен пример поведения автокорреляционной функции на чистом тоновом кадре и на плохо обусловленном. Из рис. 2 видно, что вытекает потребность к переходу к непокадровой обработке сигнала, в которой будет присутствовать постоянное уточнение тонового участка.

При автоматическом вычислении машина ищет наибольшую вершину от точки отсчета, после чего считает количество отсчетов между этими точками. В случае как на рис. 2б, произойдет ошибка – период основного тона вычислится неверно, так как кадр является шумовым, и тональная периодичность отсутствует. На подобном кадре могут содержаться и тоновые составляющие, которые при обычной линейной обработке будут безвозвратно утеряны. На данный момент эффективного фрагментатора слитной речи на звуки не существует.

Одна из причин, затрудняющая решение этой задачи, заключается в том, что не создано эффективных прогнозистов ожидаемой длины очередного звука слитной речи. Для того чтобы создать прогнозиста времени звуков, необходимо знать

статистику их распределения. Простейшим способом получения этой статистики является использование классического линейного предсказателя «тон/шума» [1].

На рис. 3 приведена блок-схема алгоритма классификатора тон/шум. Сигнал проходит обработку на линейном предсказателе, выдает ошибку и значение периода основного тона. Впоследствии ошибка проходит на пороговый классификатор и заносится в базу шумов, на которой учится нейросетевой корректор «тон/шума». Постоянное поступление сигнала способствует накоплению базы, что дает постоянное уточнение порогов за счет непрерывного обучения нейронной сети.

Континуально-дискретная модель распределения значений ожидаемой длины последовательностей тональных звуков слитной речи

Был проделан следующий эксперимент, состоящий в анализе 30-минутного речевого аудиофайла, прочитанного одним диктором. Далее зву-

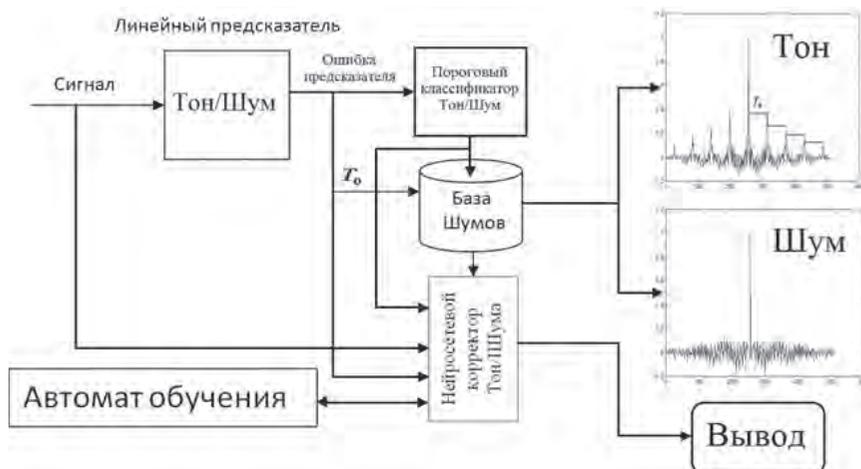


Рис. 3. Алгоритм классификатора «тон/шум»

ковой файл был разбит на тональные звуковые фрагменты. Гистограмма распределений интервалов длительности тональных фрагментов звука приведена на рис. 4.

На рис. 4 дана аппроксимация гистограммы смесью из шести нормальных законов распределений значений длин интервалов. Аппроксимация осуществлена, опираясь на зрительное восприятие. Мы видим, что наиболее вероятное значения длин одного, двух, трех и т.д. звуков кратно друг другу. Это условие может быть записано следующей системой уравнений

$$\begin{cases} E(T_1) = C_{Бuo} ; \\ E(T_2) = 2C_{Бuo} ; \\ \dots\dots\dots \\ E(T_n) = nC_{Бuo} , \end{cases} \quad (1)$$

где $E(T_n)$ – математическое ожидание нормальных распределений. Очевидно, что с увеличением статистики распределений точность определения средней длины звука будет увеличиваться.

На рис. 5 приведена гистограмма распределений, полученных при 30-минутном непрерывном чтении текста женщиной. Из рис. 5 видно, что средняя длина звука несколько другая, но так же сохраняется и есть возможность ее выделения.

Статистическая модель предсказателя длины звуков существенно упрощается, если принять гипотезу одинаковых значений среднеквадратических отклонений распределений длин звуков:

$$\sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma_{Бuo} . \quad (2)$$

Еще одним параметром статистической модели является отношение значений максимумов нормальных распределений. Из рис. 4 видно, что наибольшим является первое распределение, а

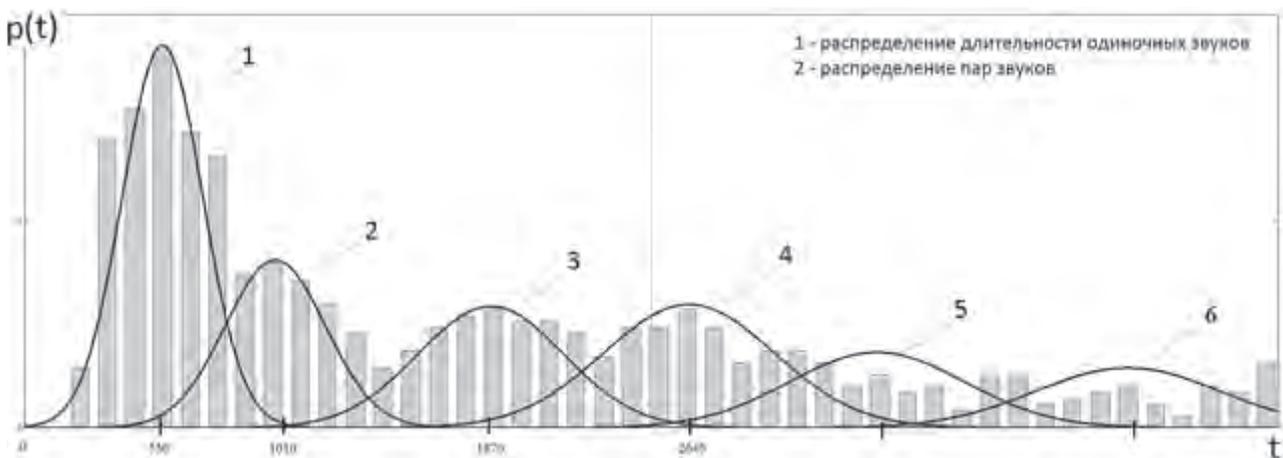


Рис. 4. Смесь нормальных законов распределения значений длительности звуков тональной слитной мужской речи

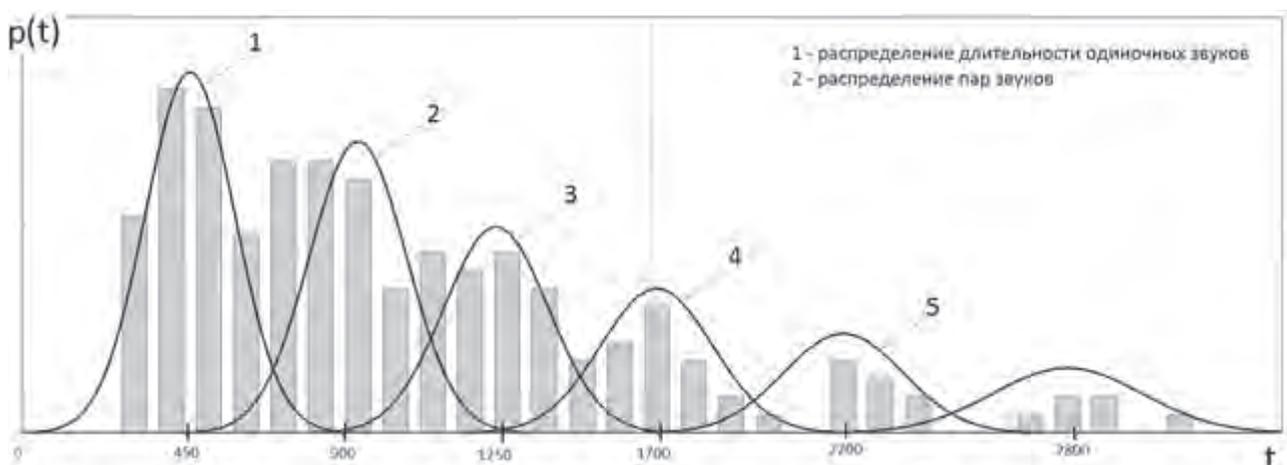


Рис. 5. Смесь нормальных законов распределения значений длительности звуков тональной слитной женской речи

последующее уменьшается. Значения P_1, P_2, \dots, P_n монотонно убывают, однако

$$P_1 + P_2 + \dots + P_n \approx 1. \quad (3)$$

Последнее обусловлено тем, что каждое из нормальных распределений появляется со своей вероятностью. Сумма всех вероятностей событий должна быть единичной. Оценить вероятность появления каждого из шести нормальных законов можно путем выделения равных интервалов вокруг математических ожиданий и подсчета попавших в эти интервалы значений:

$$P_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{E(T_i)-\sigma}^{E(T_i)+\sigma} e^{-\frac{(E(T_i)-t)^2}{2\sigma^2}} dt. \quad (4)$$

В (4) принят интервал размером в два среднеквадратических отклонения от математического ожидания. Вычисления (4) могут быть проведены однократно, так как отношения рассматриваемых вероятностей не зависят от диктора, являясь параметрами языка, на котором говорит диктор:

$$\frac{P_i}{P_1} = k_{1,i}; \quad \frac{P_i}{P_2} = k_{2,i}; \quad \frac{P_i}{P_n} = k_{3,i}. \quad (5)$$

Суперпозиция нормальных законов распределения может быть представлена как

$$p_1\varphi_1 + p_2\varphi_2 + p_3\varphi_3 + \dots + p_n\varphi_n \approx 1, \quad (6)$$

где $\varphi_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(E(T_i)-t)^2}{2\sigma^2}}$; n – число распределений.

Таким образом, опираясь на вышеописанную статистическую модель, можно создать достаточно эффективные предсказатели длин звуков слитной речи. Эта модель должна постоянно подстраиваться по своим параметрам при анализе речи. По мнению автора, человек способен надежно предсказывать интервалы звуков, опираясь на 2-3 слова, произнесенные диктором. Искусственный автомат, предсказывающий длину звука, должен работать, опираясь на статистику порядка 20-30 слов, произнесенных диктором.

Вопрос определения шума как в отдельности и накапливании базы образов шумов является не менее важным. Опираясь на его знание, с увеличением статистики точность определения границ тона становится выше. При произношении слов диктором могут возникнуть помехи со стороны, что повлияет на участки звука, и из-за этого автомат может посчитать это за тон, что скажется на точности вычисления математического ожидания периода основного тона.

Заключение

Суть проводимых исследований заключается в построении автомата по вычислению средней длины звука и его привязке к классификатору «тон/шум». Их взаимная работа будет способна уточнять параметры и накапливать базу по мере поступления сигнала.

Литература

1. Грей М. Линейное предсказание речи. М.: Связь, 1980. – 154 с.
2. Иванов А.И. Идентификация человека по особенностям его голоса. Пенза: ПНИЭИ, 2011. – С. 25-28
3. Иванов А.И. Подсознание искусственного интеллекта: программирование автоматов нейросетевой биометрии языком их обучения. Пенза: ПНИЭИ, 2012. – 125 с.
4. Волчихин В.И., Иванов А.И., Фунтиков В.А., Малыгина Е.А. Перспективы использования искусственных нейронных сетей с многоуровневыми квантователями в технологии биометрико-нейросетевой аутентификации // Известия вузов. Поволжский регион. Технические науки. №4(28), 2013. – С. 88-99.
5. Rabiner L.R., Jang B.H. Fundamentals of speech recognition. New Jersey: Prentice Hall PTR, Englewood Cliffs, 1993. – 507 с.
6. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. М. Радио и связь, 2000. – 456 с.
7. Гришин В.М., Калашников Д.М. Речевой фрагментатор для нейросетевого биометрического вокодера // Безопасность информационных технологий. №8, 2012. – С. 73-76.

BIOMETRICS: A PREDICTION OF DURATION OF VOICE-FREQUENCY SOUNDS OF CONJOINT SPEECH

Kalashnikov D.M.

The article considered the problem of division the conjoint speech into sounds, definition of their parameters and formation the bases of sounds characteristic for a voice of the individual person. It is offered to build

a distribution of lengths of voice-frequency sounds and to predict the most probable length of one sound of conjoint speech on it.

Keywords: autocorrelation function, linear speech prediction, without frame processing of a signal, mix of normal laws of distribution, fragmentator of conjoint speech.

Калашников Дмитрий Михайлович, аспирант Кафедры высшей и прикладной математики Пензенского государственного университета, инженер-программист ОАО «Пензенский научно-исследовательский электротехнический институт». Тел. 8-960-329-82-45; E-mail: riderofthesun@gmail.com

УДК 621.396.677; 621.397.671

МОДУЛЬ-РЕТРАНСЛЯТОР ДЛЯ ИНФОРМАЦИОННОЙ ЗАЩИТЫ СЛУЧАЙНЫХ АНТЕНН

Маслов О.Н.

Представлены результаты моделирования характеристик модуля-ретранслятора (МР) сигнала, содержащего конфиденциальную информацию (КИ) коммерческого назначения. В системах активной защиты КИ от утечки через случайные антенны МР является источником преднамеренной имитирующей помехи.

Ключевые слова: конфиденциальная информация, активная защита, случайные антенны, имитирующая помеха, модуль-ретранслятор.

Введение

Принципы проектирования систем активной защиты (САЗ) конфиденциальной информации (КИ) коммерческого назначения от утечки во внешнюю среду через случайные антенны (СА) различного вида (далее для краткости САЗ СА) рассмотрены в [1-2]. В [3-4] показано, что наиболее предпочтительным является применение в САЗ СА низкоэнергетических имитирующих помех (ИП), отличающихся от КИ-сигнала лишь отсутствием подлежащего защите контента (информационного содержания). Важное самостоятельное значение имеет защита КИ от утечки через апертурные СА различной конфигурации [5-7]. Предложенные технические решения [8-10 и др.] для реализации разных вариантов САЗ СА нуждаются в конструкторской проработке, макетировании и экспериментальной проверке.

В условиях вуза эффективным способом апробирования и внедрения научных результатов в сферу информационной безопасности является подготовка инновационных по содержанию лабораторно-практических занятий с будущими бакалаврами и магистрами телекоммуникаций. С одной стороны, это придает конкретику и практический смысл проводимым теоретическим исследованиям, с другой стороны, связано с необходи-

мостью наглядной и убедительной демонстрации работы САЗ СА при достаточно скромных возможностях имеющихся инструментально-измерительных средств. Выходом из ситуации может быть разработка нового оборудования, адаптированного к условиям эксплуатации в учебно-исследовательской лаборатории вуза.

Цель статьи – анализ результатов разработки, реализации и испытания одного из ключевых компонентов низкоэнергетической САЗ – модуля-ретранслятора (МР) со встроенным нелинейным элементом (НЭ) [1; 3] для генерации имитирующей помехи (ИП) с пониженным уровнем мощности, предназначенного для защиты компьютерных систем и сетей от утечки КИ через СА в окружающую среду, используемого в учебном процессе.

Эффективность МР, размещенного в открытом пространстве

Рассмотрим случай, когда МР возбуждается сосредоточенным источником КИ-сигнала, удаленным от него в свободном пространстве на расстояние r_1 . Геометрию решаемой задачи иллюстрирует рис. 1: источник КИ-сигнала расположен в точке O ; МР, показанный утолщенной линией – в точке R ; точка наблюдения M удалена от МР на расстояние r_2 . Расстояние OM от источника КИ-сигнала до точки наблюдения равно $r_{12} = \sqrt{r_1^2 + 2r_1 r_2 \cos \theta_2 + r_2^2}$, где угол θ_2 показан на рис. 1. Плотность потока мощности (ППМ) КИ-сигнала, воздействующего на МР, равняется $P_{C1} = P_C G_C / 4\pi r_1^2$, где P_C и G_C – подводимая мощность и коэффициент усиления антенны источника КИ-сигнала. В точке приема M , удаленной от источника КИ-сигнала на расстояние r_{12} , аналогичным образом имеет место $P_{C2} = P_C G_C / 4\pi r_{12}^2$.