

Обнаружение пауз в речевых сигналах

В.А. Волченков

Обработка речевой информации является на сегодняшний момент достаточно актуальной задачей и находит применение в различных сферах инженерной деятельности [1,2]. В данной статье, хотелось бы рассмотреть тему обнаружения пауз в речевых сигналах.

Длительность пауз в речи колеблется в широких пределах и может достигать до 3 с и более, имеет случайный характер. Но все же вероятность появления пауз длительностью свыше 2 с мала. Как правило, паузы разделяют на короткие (примерно до 40 мс) между элементами речи (например, на смычных звуках) и более длинные, обусловленные смысловым содержанием речи и ее ритмической структурой [4]. До сих пор в системах сжатия речевой информации и системах связи практический интерес представляло статистическое распределение вторых.

Однако, если проанализировать общую статистику для пауз, то наибольшей плотность вероятности распределения (ПВР) пауз по длительности оказывается в интервале 5...50 мс и вне этого интервала быстро убывает [3].

Очевидно, интерес только к длинным паузам был вызван подходом к реализации устройств, их обнаруживающих. Реализация существующих алгоритмов обнаружения пауз базируется на предположении, что речь – нестационарный сигнал, форма спектра речи изменяется обычно через короткие отрезки времени (около 20-30 мс). Фоновый шум считают стационарным на более длинном отрезке времени. Уровень фонового шума находится ниже уровня речевого сигнала [5]. Речь обычно делят на отрезки длительностью 16-32 мс, и анализируют уровень энергии сигнала на каждом интервале, а также количество переходов сигнала через ноль. В том случае, когда временной интервал определяется обнаружителем как пауза, перед окончательным принятием решения, что сигнал отсутствует, системе необходимо последовательно продетектировать ещё несколько фреймов (в системе GSM 5–6). Таким образом, существующие на сегодняшний день способы

определения активности речи позволяют выявить паузы, длительность которых значительно превышает 40 мс.

Для проведения эксперимента была выбрана тестовая фраза: «Продолжение отладки устройства». На рис. 1 представлена осциллограмма данной фразы и результаты обработки отрезка речи для указанной выше фразы одним из существующих детекторов активности речи, осуществляющих разделение на активные участки речи и паузы посредством разделения входного акустического сигнала на интервалы по 20 мс и сравнения кратковременной энергии каждого окна с предварительно вычисленным пороговым значением [6,7]. Разделение на активные участки речи и паузы начинается при длине пауз больше 60 мс, первые 40 мс паузы детектируются как речь.

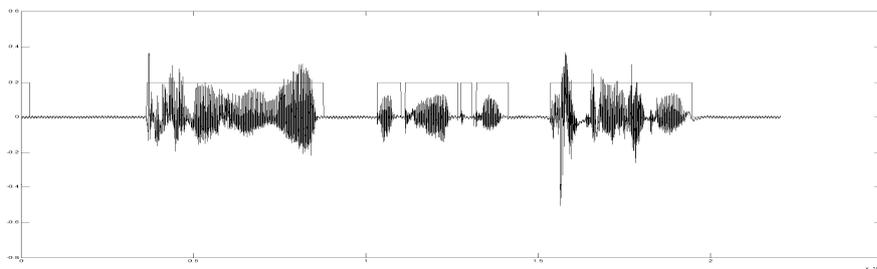


Рис. 1. Осциллограмма тестовой фразы и результаты обработки речевого сигнала одним из существующих детекторов активности речи

Таким образом, обнаружение коротких пауз и установление более точных границ для длинных пауз становятся важными задачами, решению которых и посвящена данная работа.

В настоящей работе предложен обнаружитель пауз в речевых сигналах, который обеспечивает «существенное повышение вероятности правильного разделения речевых сигналов на периоды активной речи и паузы» [10].

Структурная схема предлагаемого детектора изображена на рис. 2.

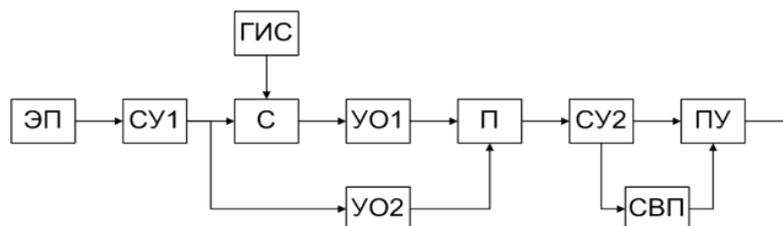


Рис. 2. Структурная схема предлагаемого детектора активности речи
Принцип работы предлагаемого детектора подробно описан в [8-10].

На рис. 3а представлена осциллограмма данной фразы и результаты обработки речевого сигнала для указанной последовательности слов.

Общее время записи сигнала – 5 с, а суммарное время активной речи составило 2,21 с (44,2 %). Правильное детектирование активных участков речи и пауз начинается при отношении сигнал-шум 7-10 дБ.

На рис. 3б показано определение пауз в начале слова «отладки». Пауза между звуками «о» и «т» 60 мс. Пауза между звуками «т» и «л» 6 мс.

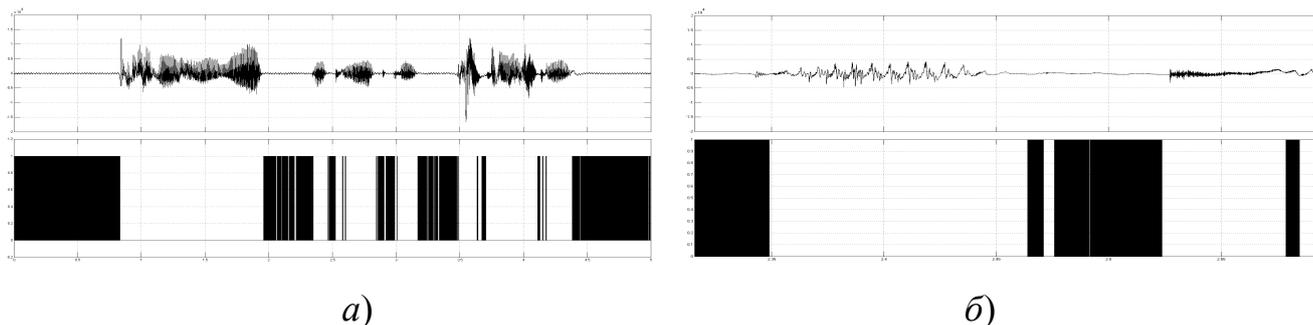


Рис. 3. Осциллограмма тестовой фразы и сигнала с выхода обнаружителя (а), определение пауз в начале слова «отладки» (б)

Был проведен эксперимент, где в интервалы, в которых находились паузы, записывался комфортный шум, параметры которого соответствовали параметрам шума, присутствовавшего в исходной записи. Качество полученного речевого сигнала практически не отличалось от исходного. Слова были хорошо различимы, речь легко воспринималась на слух.

Таким образом, в заключении можно сделать вывод, что разработан новый способ обнаружения пауз в речи, который позволяет существенно повысить точность разделения на активные участки речи и паузы. При этом качество восстанавливаемой речи остается практически на том же уровне.

Литература:

1. Астапов К.А. Применение вейвлет-преобразования для сокращения области значения искусственных нейронных сетей на примере задачи распознавания речи [Электронный ресурс] // «Инженерный вестник Дона», 2009, №1. – Режим доступа: <http://www.ivdon.ru/magazine/archive/n1y2009/105> (доступ свободный) – Загл. с экрана. – Яз. рус.

2. Марьев А.А. Метод интерпретации результатов измерений параметров речевого сигнала в задачах диагностики психоэмоционального состояния человека по его речи [Электронный ресурс] // «Инженерный вестник Дона», 2011, №4. – Режим доступа: <http://www.ivdon.ru/magazine/archive/n4y2011/538> (доступ свободный) – Загл. с экрана. – Яз. рус.

3. Вахитов Ш.Я. Акустика: Учебник для вузов [Текст] / Ш.Я. Вахитов., Ю.А. Ковалгин, А.А. Фадеев, Ю.П. Щевьев; Под ред. профессора Ю.А. Ковалгина. – М.: Горячая линия–Телеком, 2009. – 660 с.: ил.

4. Михайлов В.Г. Измерение параметров речи [Текст] / В.Г. Михайлов, Л.В. Златоустова; Под ред. М.А. Сапожкова. – М.: Радио и связь, 1987. – 168 с.: ил.

5. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи [Текст] / Под ред. О.И. Шелухина. — М.: Радио и связь, 2000. — 456 с.

6. Sohn J. A voice activity detector employing soft decision based noise spectrum adaptation [Текст] / J. Sohn and W. Sung // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing / Seattle, WA, 1998. – Vol. 1. – pp. 365-368.

7. Kondoz A.M. Digital Speech. Coding for Low Bit Rate Communication Systems. [Текст] – John Wiley & Sons, Ltd. 2004. – 442 p.

8. Пат. 2436173 Российская Федерация, МПК G10L 15/00, G10L 11/02, Способ обнаружения пауз в речевых сигналах и устройство его реализующее [Текст] / Витязев В.В., Розов В.И., Волченков В.А.; заявитель и патентообладатель Рязанский государственный радиотехнический университет. – № 2010124342/08, заяв. 15.06.10; опубл. 10.12.11, Бюл. 34.

9. Волченков В.А. Методы и алгоритмы детектирования активности речи [Текст] / Волченков В.А., Витязев В.В. // Цифровая обработка сигналов. 2013. №1. С. 54–60.

10. Волченков В.А. Детектор активности речи [Текст] / Волченков В.А., Витязев В.В. // Труды РНТОРЭС им. А.С. Попова. Серия: Цифровая обработка сигналов и её применение. Выпуск: XIII – 2. / Москва: РНТОРЭС им. А.С. Попова, 2011. С. 256 – 258.