Математические структуры и моделирование
2011, вып. 22, с. 24–27

УДК 519.688:534.4

АЛГОРИТМ ВЫДЕЛЕНИЯ ОСНОВНОГО ТОНА И ДЕТЕКТИРОВАНИЯ ТОН/НЕ ТОН ПО МИНИМУМАМ РАЗНОСТНОЙ ФУНКЦИИ НА УЧАСТКЕ МИНИМАЛЬНОГО ПЕРИОДА

Е. А. Первушин, Д. Н. Лавров

Описывается алгоритм нахождения мгновенных значений периодов основного тона речевого сигнала, использующий кратковременную функцию среднего значения разности. Решение о наличии тона принимается при сравнении значений минимумов разностной функции. Предлагается альтернативный выбор начала периода.

Процедура выделения основного тона является одной из важнейших задач в области анализа речи. Выделение основного тона используется в вокодерах, системах синтеза речи, системах распознавания по голосу и других приложениях. В речевом сигнале период основного тона соответствует периоду колебаний голосовых связок и является одной из основных характеристик источника возбуждения голосового тракта.

Существует ряд алгоритмов, каждый из которых имеет свои преимущества и недостатки. Пиковые алгоритмы выделения основного тона используют амплитудно-временные характеристики сигнала, выделяя точки локальных максимумов и/или минимумов речевого сигнала и используя их для определения периода. Данные методы чувствительны к появлению ложных максимумов.

Суть спектрального метода заключается в построении спектра сигнала и нахождении максимума в пределах допустимых частот. При этом для вычисления требуется относительно большое количество операций.

Алгоритмы, использующие автокорреляционную или кратковременную функцию среднего значения разности (КФСР) [4], вычисляют некоторую интегральную для заданного интервала характеристику и по ней оценивают значение периода. При этом вычисленное значение при соответствующем нормировании является также мерой вокализованности. Алгоритмы данного класса обладают приемлемым сочетанием простоты и точности, однако чувствительны к изменениям формы речевого сигнала. Поэтому в данной работе предлагается модификация алгоритма, основанного на вычислении КФСР. Выбор между

Омский государственный университет им. Ф. М. Достоевского.
E-mail: pervushin_evgen@mail.ru

автокорреляционной функцией и КФСР основан на более простом вычислении последней. Более детальное сравнение этих функций можно найти в [2].

В общем виде КФСР (иногда эту функцию называют разностной или сдвиговой) определяется как

$$S_n(h) = \sum_{m=-\infty}^{\infty} |s(n+m)w_1(m) - s(n+m-h)w_2(m-h)|,$$

где $s(t)$ — функция сигнала; w_1, w_2 — оконные функции. Очевидно, что если участок $s(t)$, попадающий в окно, имеет квазипериодический характер с периодом T, то функция S_n будет иметь ярко выраженные минимумы при $h = T, 2T, \ldots$ Условие наличия такого минимума будет использоваться для определения тон/не тон, т. е. является ли участок вокализованным или невокализованным. Пример вокализованных и невокализованных участков речи, а также соответствующих им функций разности, приведён на рис. 1.

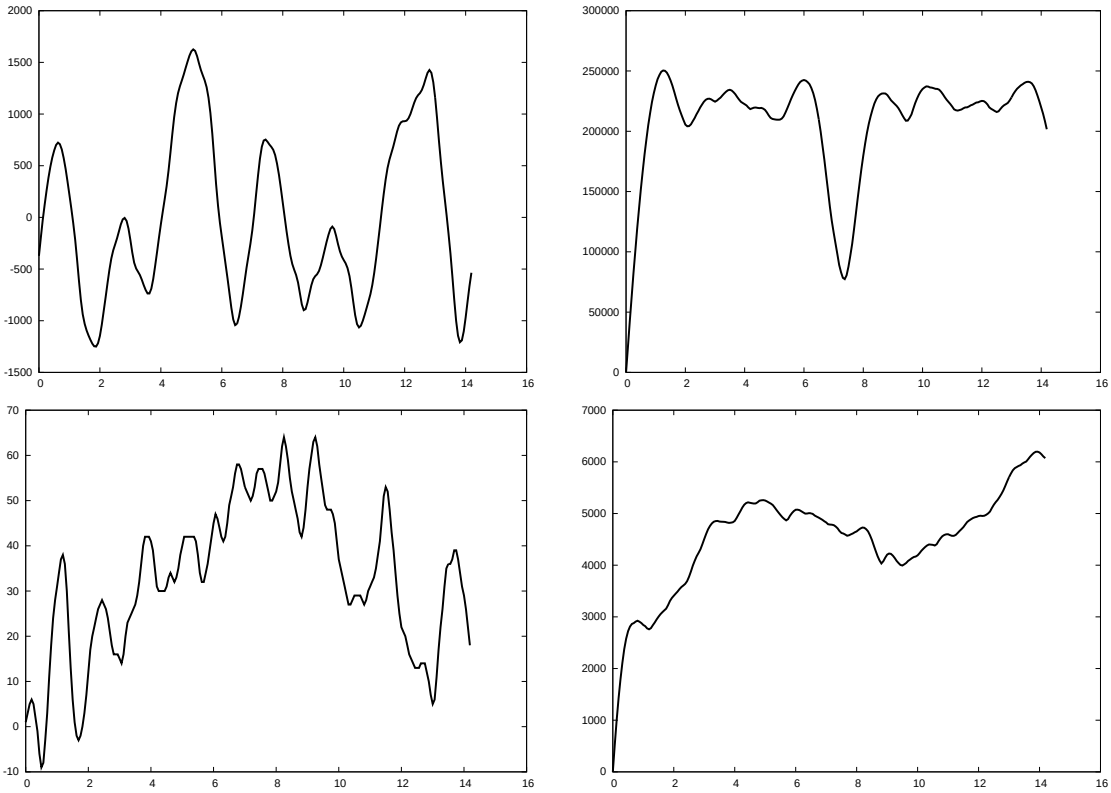

Рис. 1. Пример участков сигналов и их разностных функций: вокализованный участок (сверху), невокализованный (снизу)

Для работы алгоритма потребуется зафиксировать несколько констант:

T_{max} — максимальное значение периода, выраженное в отсчётах;

(P_1, P_2) — интервал, использующийся для проверки на наличие тона;

A_1 — коэффициент для задания порога на наличие тона;

A_2 — коэффициент для задания порога, использующийся в проверке, не найден ли период, кратный основному.

На каждой итерации алгоритма исследуется часть сигнала, попадающая в прямоугольное окно размера T_{max}

$$w_1(t) = w_2(t) = \begin{cases} 1, & 0 \le t \le T_{max} \\ 0, & \text{в противном случае} \end{cases}$$

1. Пусть $s_i,\ i = 1, \ldots, T_{max}$ — отсчёты сигнала, попадающие в окно на данной итерации.
2. Вычисляются значения разностей

$$S_h = \sum_{i=1}^{T_{max}} |s_i - s_{i+h}|,\ h = P_1, \ldots, T_{max}.$$

3. Вычисляется наиболее вероятное значение периода, выраженное в отсчётах

$$T = \arg \min_{h \in (P_1, T_{max})} S_h.$$

4. Если $T < P_2$, то полагаем, что на данном участке нет основного тона; конец итерации.
5. Если существует i такой, что

$$S_i \le A_1 S_T,\ i \in (P_1, P_2),$$

также полагаем, что на данном участке нет основного тона, и переходим к следующей итерации.
6. Обозначим $S = S_T$.
7. Находим минимум среди значений $S_{P_2}, \ldots, S_{T-P_1}$. Пусть T_1 — точка, в которой достигается минимум. Если $S_{T_1} \le A_2 S$, то положим $T = T_1$ и вернёмся к началу шага 7.
8. Считаем полученное значение T периодом основного тона.

В случае нахождения периода, окно сдвигается на величину T, в противном случае — на величину $T_{max}/2$.

Константы, необходимые для работы алгоритма, определим следующим образом. Положим, что алгоритм должен искать периоды основного тона, соответствующие диапазону частот F_{0min} — F_{0max} Гц, тогда $T_{max} = 1/F_{0min} rate$, где $rate$ — частота дискретизации сигнала. Определим теперь $T_{min} = 1/F_{0max} rate$ — минимальное значение периода основного тона. Положим $P_1 = 1/3T_{min}$, $P_2 = 2/3T_{min}$, тогда (P_1, P_2) — подынтервал интервала $(0, T_{min})$, в котором маловероятно появление минимума значений S_h. При фиксированных значениях P_1 и P_2 настраиваются параметры A_1, A_2 для удовлетворения целей приложения, в котором используется данный алгоритм.

Так, например, для системы идентификации дикторов, предложенной в [3], алгоритм выделения основного тона должен быть настроен таким образом, чтобы, с одной стороны, он не выделял периоды, на которых нельзя с уверенностью утверждать наличие тона, и с другой стороны, должно быть выделено достаточное количество периодов для представления шаблона диктора. Эффективность

выбранных значений параметров в данном приложении оценивается по итоговому проценту верных идентификаций при данной базе. При заданных значениях $F_{0min} = 70$ Гц, $F_{0max} = 450$ Гц в результате экспериментов над тестовой базой были выбраны значения порогов $A_1 = 1,3$ и $A_2 = 1,15$.

Помимо определения мгновенных значений периодов основного тона, часто требуется указать начало периода, например, для системы синтеза речи. Распространёнными являются следующие два подхода. Один из них в качестве начала периода определяет точку максимального значения сигнала в пределах найденного периода. Такой подход проще в реализации и, возможно, способен на более точное определение начала периода ввиду определённой выраженности максимума речевого сигнала.

Однако более часто, особенно в системах синтеза (см., например, [1]), используется другой подход, который определяет в качестве начала периода точку перехода сигнала через ноль слева от точки максимума. В своей работе автор использует и другие подходы, при которых сначала находится точка максимума, а затем определяется начало периода, отстоящее слева от точки максимума на расстояние, фиксированное по времени либо зависящее от длины найденного периода.

Предложенный в данной работе алгоритм выделения основного тона, принятия решения тон/не тон и разметки на периоды использовался в системе распознавания дикторов. Алгоритм не требует сложных вычислений и многочисленных настроек и может быть применён в системах анализа и синтеза речи, распознавании речи, распознавании дикторов и других приложениях.

Литература

1. Бабкин А. В. Автоматический синтез речи — проблемы и методы генерации речевого сигнала // Труды Международного семинара по компьютерной лингвистике и её приложениям «Диалог'98». М., 1998.
2. Баронин С. П. Автокорреляционный метод выделения основного тона речи. Пятьдесят лет спустя // Речевые технологии. 2008. Вып. 2. С. 3–12.
3. Первушин Е. А., Лавров Д. Н. Система идентификации диктора на основе выделения информативных участков речевого сигнала // Материалы II межвузовской научно-практической конференции ОмГТУ «Информационные технологии и автоматизация управления». 2010. С. 188–189.
4. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов: пер. с англ. М., 1981. 496 с.