

МОДИФИЦИРОВАННАЯ ПОЛЮСНАЯ МОДЕЛЬ РЕЧЕВОГО СИГНАЛА
И ЕЕ ПРИМЕНЕНИЕ ДЛЯ ПОДАВЛЕНИЯ ШУМА

ЛЮДОВИК ЕВГЕНИЙ КУЗЬМИЧ

Институт кибернетики им. В.М.Глушкова АН УССР
Киев, СССР 252207

АННОТАЦИЯ

Дополнение традиционной полюсной модели (модели линейного предсказания) речевого сигнала моделью квазипериодического сигнала возбуждения, не критичной к наиболее частым ошибкам в выделении основного тона, позволяет использовать периодичность вокализированной речи для ее выделения из смеси с шумом.

ВВЕДЕНИЕ

Традиционная полюсная модель отражает квазипериодичность сигнала возбуждения на вокализированных интервалах сомножителем $(1 - z^{-L})$, стоящим в знаменателе передаточной функции речевого тракта. Таким образом, сигнал возбуждения считается строго периодическим с периодом основного тона L . Известно, однако, что свойство периодичности наблюдается в реальном сигнале лишь в определенной мере, иногда больше, иногда меньше. С одной стороны, форма сигнала изменяется от периода к периоду, с другой — изменяется и сам период.

Вследствие такого рода отличий реального квазипериодического сигнала от модельного строго периодического в процессе анализа возникают ошибки в определении основного тона. Все ошибки могут быть разбиты на три класса:

- 1) "малые" ошибки в пределах 10–15% от истинного значения периода;
- 2) ошибочные значения, кратные периоду или частоте основного тона;

- 3) грубые ошибки, не коррелирующие с истинным значением периода.

Помимо указанных проблем на этапе анализа неадекватность строго периодической модели вызывает некоторую неестественность синтезируемого вокализированного сигнала.

Попытки /1/ использования на основе этой жесткой модели свойства квазипериодичности для коррекции вокализированных речевых сигналов, искаженных аддитивным шумом, наталкиваются на следующие трудности:

- 1) необходимо точно определять период основного тона и признак тон/шум по зашумленному сигналу, что и в отсутствие шума является сложной задачей;
- 2) непонятно, как устанавливать соответствие между отсчетами сигнала из разных периодов и с какими весами следует их усреднять.

Навязывание строгой периодичности с зачастую ошибочным периодом приводит к "смазыванию" динамики сигнала и снижению разборчивости.

Таким образом, имеется потребность в модели квазипериодичности, которая отражала бы изменчивость сигнала от периода к периоду, изменение самого периода, а также была бы не критична к ошибкам в определении периода основного тона.

МОДЕЛЬ

Первый шаг в направлении такой модели можно усмотреть в работе /2/, в которой выражение $(1 - z^{-L})$ заменяется на

$(1 + \varepsilon_{L-1}z^{-(L-1)} + \varepsilon_L z^{-L})$, а параметры ε_{L-1} , ε_L и L определяются по речевому сигналу. В /2/, однако, эта замена использована фактически только для усовершенствования корреляционного метода выделения основного тона и уточнения амплитуды периодического сигнала возбуждения. На самом же деле такой вариант позволяет учесть в модели изменчивость периода и формы сигнала, но не снимает проблемы создания модели, не критичной к ошибкам.

В настоящей работе предлагается заменить выражение $(1 - z^{-L})$ на выражение $(1 + \sum_{i=0}^{L-1} \varepsilon_{L-i} z^{-(L-i)} + \sum_{i=0}^{2L-1} \varepsilon_{2L-i} z^{-(2L-i)})$, причем параметры ε и L должны определяться по речевому сигналу.

Во временной области предлагаемая модель голосового источника имеет вид:

$$w_n = -\sum_{i=0}^{L-1} \varepsilon_{L-i} w_{n-L+i} - \sum_{i=0}^{2L-1} \varepsilon_{2L-i} w_{n-2L+i} + e_n \quad (1)$$

где e_n — входной сигнал типа белого шума,
 w_n — квазипериодический сигнал голосового возбуждения.

Поскольку в предлагаемой модели учитывается периодичность и с периодом L и с периодом $2L$, вероятность грубых ошибок, некоррелированных с истинным периодом, снижается. Кроме того, здесь малосущественно, равен ли период L или $2L$. Таким образом, модель не критична по отношению к довольно частым ошибкам типа удвоения периода или удвоения частоты основного тона.

Результирующая модифицированная полюсная модель получается, если квазипериодический сигнал w_n подать на вход обычной линейной прогнозирующей модели:

$$x_n = -\sum_{i=1}^m a_i x_{n-i} + w_n \quad (2)$$

Передаточная функция модифицированной полюсной модели имеет вид:

$$H(z) = \frac{1}{\sum_{i=0}^m a_i z^{-i}} \times \frac{1}{(1 + \sum_{i=0}^{L-1} \varepsilon_{L-i} z^{-(L-i)} + \sum_{i=0}^{2L-1} \varepsilon_{2L-i} z^{-(2L-i)})} \quad (3)$$

ИДЕНТИФИКАЦИЯ МОДЕЛИ

Задача оценивания параметров модели по отрезку речевого сигнала x_n на основе метода максимального правдоподобия или метода линейного предсказания сводится к минимизации следующего критерия:

$$F(a, \varepsilon, L) = \sum_{i,j} R_{i-j} a_i a_j + \sum_s \varepsilon_s \varepsilon_{s+i} \quad (4)$$

где $R_i = \sum_n x_n x_{n+i}$, $a_0 = \varepsilon_0 = 1$,

$$\varepsilon_i = 0, \quad i \notin \{0, L-1, L, 2L-2, 2L-1, 2L\}$$

Для решения этой задачи предлагается итерационный алгоритм, каждая итерация которого состоит из двух этапов.

На первом этапе при фиксированных значениях параметров модели голосового источника (на начальной итерации полагаем $\varepsilon_i = 0, i \neq 0$) осуществляется минимизация по параметрам a , определяющим резонансные свойства речевого тракта, что сводится к решению традиционной для метода линейного предсказания системы уравнений:

$$\sum_{j=1}^m R_{i-j} a_j = -R_i^{\varepsilon}, \quad 1 \leq i \leq m, \quad (5)$$

где $R_i^{\varepsilon} = \sum_j R_{i-j} \sum_s \varepsilon_s \varepsilon_{s+j}$ — корреляционная функция сигнала x_n , из которого путем обратной фильтрации устранена информация об основном тоне.

На втором этапе при фиксированных значениях параметров a осуществляется минимизация по параметрам модели голосового источника ε и L . При этом для каждого возможного значения L находится минимум критерия (4) по параметрам ε что сводится к решению следующей системы

уравнений:

$$\sum_{j=L-1}^L R_{i-j}^a \mathcal{G}_j + \sum_{j=2L-2}^{2L} R_{i-j}^a \mathcal{G}_j = -R_i^a, \quad (6)$$

$$i = L-1, L, 2L-2, 2L-1, 2L,$$

где $R_i^a = \sum_j R_{i-j} \sum_r a_r a_{r+j}$ - корреляционная функция сигнала, из которого путем обратной фильтрации устранена формантная информация.

Значение L , наилучшее для фиксированного значения параметра a , определяется путем перебора.

Поскольку на каждом этапе отыскивается глобальный минимум по соответствующей обобщенной координате, значение критерия монотонно уменьшается от итерации к итерации, или же не изменяется, если найдена точка локального минимума. Незначительное изменение критерия в результате выполнения очередной итерации и неизменность параметра L могут быть приняты за условие останова итерационного алгоритма.

ИДЕНТИФИКАЦИЯ МОДЕЛИ ПО ЗАШУМЛЕННОМУ РЕЧЕВОМУ СИГНАЛУ.

КОРРЕКЦИЯ СИГНАЛА

Рассмотрим теперь задачу коррекции вокализованного сигнала, искаженного аддитивным шумом. Пусть на полезный сигнал x_n , порожденный модифицированной полюсной моделью, наложен аддитивный шум d_n с известной спектральной плотностью, так что фактически наблюдаемым является сигнал y_n :

$$y_n = x_n + d_n.$$

Задачу выделения полезного сигнала x_n поставим как задачу отыскания максимально правдоподобных оценок сигнала x_n и параметров модели a, g, L .

После перехода в спектральную область и выполнения простых преобразований функции правдоподобия получаем критерий, подлежащий минимизации:

$$P(X, a, g, L, \sigma_e) = 2N^2 \ln \sigma_e + \sum_n |x_n|^2 |A_n|^2 |G_n|^2 / \sigma_e^2 + \sum_n |y_n - x_n|^2 / |D_n|^2, \quad (7)$$

где x_n и y_n , $0 \leq n \leq N-1$, - дискретные спектры Фурье искомого и наблюдаемого сигналов соответственно,

$|D_n|^2$ - известный энергетический дискретный спектр шума,

σ_e^2 - дисперсия сигнала возбуждения на входе модифицированной полюсной модели, подлежащая определению,

$$A_n = \sum_{s=0}^m a_s e^{-j s n 2\pi / N},$$

$$G_n = \sum_s g_s e^{-j s n 2\pi / N}, s \in \{0, L-1, L, 2L-2, 2L-1, 2L\}.$$

Особенностью выведенного критерия является наличие подстраиваемого спектра G_n , обратной спектральной характеристике голосового источника.

Для минимизации критерия предлагается итерационный алгоритм, каждая итерация которого состоит из четырех этапов, причем на каждом этапе определяется глобальный минимум по одной из четырех обобщенных переменных x_n, a, g, L и σ_e при фиксированных значениях остальных.

I-й этап. Минимизация по x_n :

$$x_n = y_n \frac{\sigma_e^2}{|A_n|^2 |G_n|^2} / \left(\frac{\sigma_e^2}{|A_n|^2 |G_n|^2} + |D_n|^2 \right)$$

II-й этап. Минимизация по a сводится к решению системы уравнений (5) с коэффициентами

$$R_i^g = \sum_n |x_n|^2 |G_n|^2 \cos(2\pi i n / N).$$

III-й этап. Минимизация по параметрам источника осуществляется точно так же, как в описанном выше алгоритме идентификации модифицированной полюсной модели по незашумленному сигналу, при этом

$$R_i^a = \sum_n |x_n|^2 |A_n|^2 \cos(2\pi i n / N).$$

IV-й этап. Минимизация по σ_e :

$$\sigma_e^2 = \sum_n |x_n|^2 |A_n|^2 |G_n|^2 / N^2.$$

Как и в случае первого алгоритма критерий монотонно уменьшается от итерации к итерации, либо остается неизменным, если значения обобщенных переменных соответствуют локальному минимуму. Условие останова такое же, как и в первом алгоритме.

На основании полученной оценки спектра незашумленного сигнала x_n путем обратного преобразования Фурье можно вычислить оценку отсчетов исходного речевого сигнала во временной области.

ЗАКЛЮЧЕНИЕ

Предложенная модель может быть полезна при анализе речевых сигналов, поскольку она позволяет осуществить развязку формантной информации и информации, связанной с квазипериодическим сигналом голосового возбуждения.

Достоинством модели является ее нечувствительность к ошибкам в определении периода основного тона типа удвоения периода или удвоения частоты, а также отсутствие необходимости принимать решение о призраке тон/шум.

С точки зрения задачи коррекции зашумленных вокализованных сигналов в этой модели важно то, что степень периодичности и веса при усреднении различных периодов определяются самим зашумленным речевым сигналом, в отличие от других подходов, включающих элемент произвола.

Литература

- 1/. Лим Дж.С., Опенхайм А.В. Коррекция и сжатие спектра зашумленных речевых сигналов. ТИИЭР, т. 67, № 12, 1979.
- 2/. Kwon S.Y., Goldberg A.J. An Enhanced LPC Vocoder with No Voiced/Unvoiced Switch. IEEE Transactions on ASSP, vol ASSP-32, N 4, 1984.