

УДК 004.627

О ПРИМЕНЕНИИ ИСКАЖЕННОГО ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ В КОМПРЕССИИ СИГНАЛОВ

В.О. Чернецкий

Рассмотрены некоторые особенности искаженного линейного предсказания применительно к алгоритмам компрессии речевых сигналов.

Ключевые слова: линейное предсказание, искаженное линейное предсказание, компрессия сигналов.

1. Традиционное линейное предсказание

Значительная часть методов компрессии речи основана на линейном предсказании. В этом случае оцифрованный речевой сигнал разделяется на кадры длительностью 5÷30 мс, и для каждого кадра строится модель вида

$$s[n] = \sum_{i=1}^P a_i \cdot s[n-i] + r[n], \quad (1)$$

где $s[n]$ – отсчеты оцифрованного сигнала; $r[n]$ – остатки предсказания; a_i – коэффициенты предсказания; P – порядок предсказания.

Уравнение (1) называется уравнением линейного предсказания, и ему соответствует передаточная функция

$$H(z) = \frac{s(z)}{r(z)} = \frac{1}{1 - \sum_{i=1}^P a_i \cdot z^{-i}}. \quad (2)$$

Фильтр с передаточной функцией (2) принято называть LPC-фильтром. Соответственно, фильтр с передаточной функцией $H^{-1}(z)$ называется обратным LPC-фильтром. В ряде случаев передаточную функцию LPC-фильтра представляют в виде

$$H(z) = \frac{K}{1 - \sum_{i=1}^P a_i \cdot z^{-i}}, \quad (3)$$

где $K = \sqrt{\sum_{n=1}^N r[n] \cdot r[n]}$; N – количество отсчетов в кадре.

В процессе компрессии (кодирования) речевого сигнала, в кодере с помощью обратного LPC-фильтра выделяется сигнал остатков предсказания (применительно к кодированию он называется остатками 1-го порядка). Затем для остатков 1-го порядка ищутся их некоторые параметры (набор этих параметров определяется алгоритмом компрессии). Коэффициенты

LPC-фильтра и параметры остатков квантуются и передаются в декодер. При декодировании сигнал остатков восстанавливается (в этом случае он называется сигналом возбуждения LPC-фильтра) и обрабатывается LPC-фильтром.

Порядок предсказания P определяется количеством формант в спектре речевого сигнала. При частоте дискретизации 8 КГц количество формант не превосходит четырех, поэтому минимальное значение порядка равно 9 (8 при использовании фильтра частотных предсказаний). Повышение порядка предсказания вплоть до 20 не приводит к сколько-нибудь заметному повышению детализации АЧХ фильтра. В связи с этим, в большинстве методов компрессии используется предсказание 10-го порядка (с некоторым запасом).

2. Искаженное линейное предсказание

Для повышения качества синтезированного в декодере речевого сигнала желательно улучшить детализацию АЧХ LPC-фильтра без существенного повышения его порядка, причем, в силу особенностей слухового восприятия, улучшение детализации необходимо только в низкочастотной области (на частотах ниже критической, которая равна $1 \div 2$ КГц). Это возможно, если перейти от LPC-фильтра с передаточной функцией (2) к фильтру с передаточной функцией

$$H(\tilde{z}) = \frac{1}{1 - \sum_{i=1}^P c_i \cdot \tilde{z}^{-i}}, \quad (4)$$

где $\tilde{z}^{-1} = H_0(z)$ – передаточная функция некоторого дискретного фильтра [1].

Поскольку $z^{-1} = e^{-j\omega T_0}$, где T_0 – период дискретизации, для сохранения возможности использования широко известных способов поиска параметров традиционного линейного предсказания и для обеспечения устойчивости фильтра (4), передаточная функция $H_0(z)$ должна удовлетворять следующим условиям:

- 1) фильтр $H_0(z)$ должен быть устойчивым;
- 2) $|H_0(z)| \equiv 1$, т. е. $H_0(j\omega) = e^{-j\tilde{\omega}(\omega)T_0}$;
- 3) $\tilde{\omega}(0) = 0$;
- 4) $\tilde{\omega}\left(\frac{\pi}{T_0}\right) = \frac{\pi}{T_0}$.

Данным условиям удовлетворяет дискретный фильтр 1-го порядка с передаточной функцией

$$H_0(z) = \frac{z^{-1} - \alpha}{1 - \alpha \cdot z^{-1}}, |\alpha| < 1. \quad (5)$$

В этом случае

$$\tilde{\omega}(\omega) = \frac{1}{T_0} \cdot \arccos \frac{(1 + \alpha^2) \cdot \cos(\omega T_0) - 2\alpha}{1 + \alpha^2 - 2\alpha \cdot \cos(\omega T_0)} = \omega + \frac{2}{T_0} \operatorname{arctg} \frac{\alpha \cdot \sin(\omega T_0)}{1 - \alpha \cdot \cos(\omega T_0)}, \quad (6)$$

$$\frac{d\tilde{\omega}}{d\omega} = 1 + 2\alpha \cdot \frac{\cos(\omega T_0) - \alpha}{1 + \alpha^2 - 2\alpha \cdot \cos(\omega T_0)}, \quad (7)$$

и критическая частота

$$\omega_t = \frac{1}{T_0} \cdot \arccos(\alpha). \quad (8)$$

При $0 < \alpha < 1$ и $\omega < \omega_t$, $\frac{d\tilde{\omega}}{d\omega} > 1$, что и обеспечивает повышенную по сравнению с традиционным LPC-фильтром детализацию АЧХ. При $\omega \geq \omega_t$, $\frac{d\tilde{\omega}}{d\omega} \leq 1$, и детализация АЧХ оказывается ниже, чем в случае традиционного LPC-фильтра.

Как следует из (8), значение параметра α зависит от частоты дискретизации. При этом имеется неопределенность с выбором значения критической частоты ω_t . В [1] приводится аналитическая формула для вычисления α , основанная на психоакустической модели:

$$\alpha = 1,0674 \cdot \sqrt{\frac{2}{\pi} \cdot \operatorname{arctg}(0,06583 \cdot f_0)} - 0,1916, \quad (9)$$

где f_0 – частота дискретизации (КГц).

Тем не менее, имеются исследования, которые не подтверждают корректность использования (9). Так, например, в [2] для частоты дискретизации 16 КГц экспериментально получено наилучшее качество синтезированного сигнала при $\alpha = 0,3$, тогда как, исходя из (9) в этом случае $\alpha = 0,575$. Для частоты дискретизации 8 КГц, автор данной статьи экспериментально получил $\alpha = 0,3$, хотя из (9) следует $\alpha = 0,4$.

Что касается порядка предсказания, то для проявления эффекта повышения детализации АЧХ, он, к сожалению, должен быть увеличен как минимум до 20.

3. LPC-анализ и синтез

Как и в случае традиционного линейного предсказания, LPC-анализ заключается в нахождении коэффициентов c_i передаточной функции (4) и вычислении остатков 1-го порядка для каждого кадра речевого сигнала. При использовании оконной функции Хэмминга $w[n]$, исходя из (4) с учетом (5) можно записать

$$s'[n] = \sum_{i=1}^P c_i \cdot s_i[n] + r[n], \quad (10)$$

где

$$\begin{aligned} s'[n] &= s[n] \cdot w[n], \\ s_1[n] &= -\alpha \cdot s'[n] + s'[n-1] + \alpha \cdot s_1[n-1], \\ s_i[n] &= -\alpha \cdot s_{i-1}[n] + s_{i-1}[n-1] + \alpha \cdot s_i[n-1], \\ s'[-1] &= s_i[-1] = 0, \\ n &= 0 \dots N-1. \end{aligned}$$

Минимизация энергии остатков 1-го порядка $E = \sum_{i=0}^{N-1} r[i] \cdot r[i]$ в (10) приводит к системе уравнений

$$\begin{bmatrix} R_0 & R_1 & R_2 & \dots & R_{P-1} \\ R_1 & R_0 & R_1 & \dots & R_{P-2} \\ \vdots & & & & \\ R_{P-1} & \dots & R_2 & R_1 & R_0 \end{bmatrix} \times \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_P \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_P \end{bmatrix}, \quad (11)$$

где $R_0 = \sum_{i=0}^{N-1} s'[i] \cdot s'[i]$; $R_k = \sum_{i=0}^{N-1} s'[i] \cdot s_k[i]$.

Как и в случае традиционного линейного предсказания, коэффициенты c_i непригодны для непосредственного квантования, и должны быть преобразованы в линейные спектральные пары [3]. Для облегчения их квантования могут использоваться те же приемы, что и в случае традиционного предсказания, например, добавление белого шума в речевой сигнал и демпфирование резонансов АЧХ фильтра путем умножения входящих в (11) R_k на экспоненциальную оконную функцию.

Структурная схема обратного искаженного LPC-фильтра, необходимого для выделения остатков 1-го порядка, приведена на рис. 1. Фильтр реализуем и при $|\alpha| < 1$ устойчив.

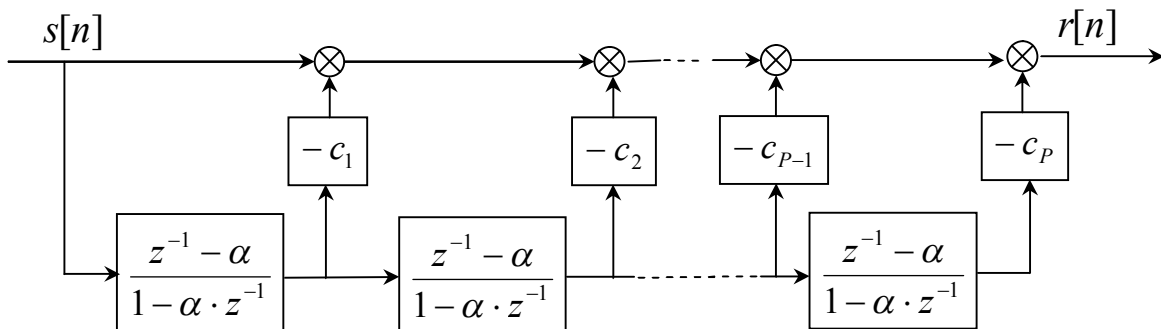


Рис. 1. Структурная схема искаженного обратного LPC-фильтра

Необходимый для синтеза речи в декодере LPC-фильтр непосредственно по передаточной функции (4) не реализуется из-за появления в структурной схеме алгебраических контуров, обусловленных наличием в числителе (5) свободного члена $-\alpha$. Развязка алгебраических контуров приводит к одной из возможных схем реализации, изображенной на рис. 2.

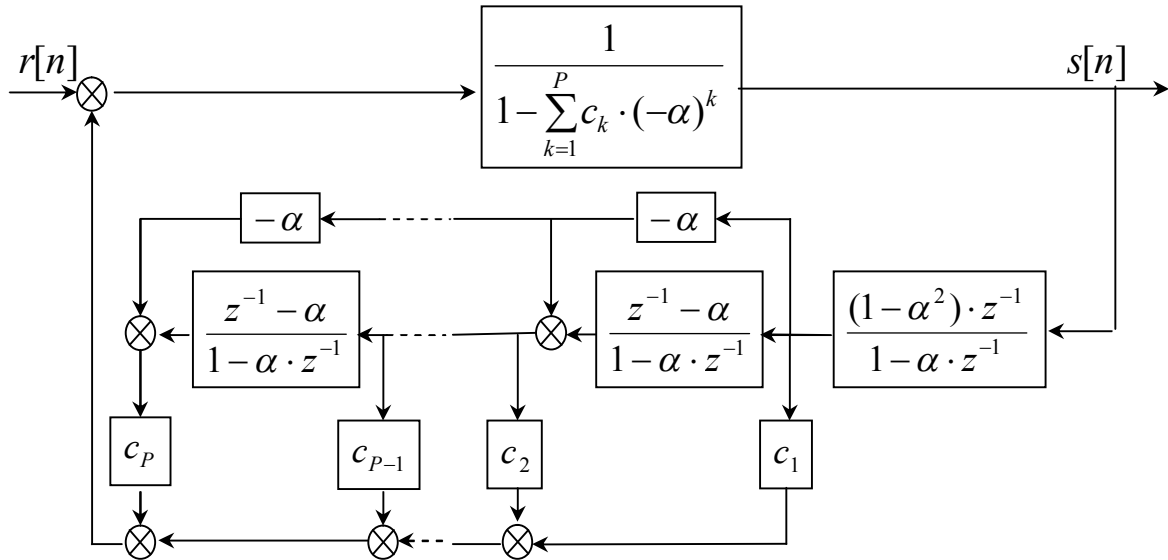


Рис. 2. Структурная схема искаженного LPC-фильтра

4. Результаты исследований

Для исследования влияния искаженного линейного предсказания на качество компрессии речи разработан и реализован CELP-алгоритм компрессии со скоростью цифрового потока 8,8 Кбит/с. Алгоритм реализован в двух вариантах – с использованием традиционного предсказания и с использованием искаженного предсказания 20-го порядка. По результатам их сравнения можно сделать следующие предварительные выводы:

1. Применение искаженного линейного предсказания приводит к заметному повышению качества синтезированной в декодере женской речи, для которой характерны небольшие значения периода основного тона. В результате, скорее всего, имеется возможность отказаться от использования дробных значений периода основного тона и связанной с этим интерполяции (повышения частоты дискретизации) при кодировании и декодировании. С точки зрения автора, качество синтезированной женской речи при использовании искаженного предсказания без интерполяции практически такое же, как и с интерполяцией, и существенно выше, чем при использовании традиционного предсказания с интерполяцией. При синтезе мужской речи с большими периодами основного тона, переход к искаженному предсказанию существенного влияния на качество не оказывает, так как в этом случае традиционное предсказание обеспечивает весьма качественный синтез даже без интерполяции.

2. Использование искаженного линейного предсказания позволяет отказаться от применения в декодере адаптивного пост-фильтра. По мнению автора, качество синтезированной речи при использовании искаженного предсказания без пост-фильтра выше, чем при использовании традиционного предсказания с пост-фильтром. Более того, включение пост-фильтра в ряде случаев приводит к снижению качества синтезированной речи. Это можно объяснить тем, что действие пост-фильтра основано на усилении формантных областей спектра. В результате маскируются дефекты синтеза, однако спектр синтезированного сигнала искажается. При использовании искаженного предсказания дефекты синтеза малозаметны, и искажение спектра пост-фильтром подавляет положительный эффект от их маскировки.

3. При использовании искаженного предсказания возможность повышения качества синтезированной речи ограничена низкой частотой дискретизации (8 КГц), поэтому наибольшего эффекта следует ожидать при компрессии широкополосных речевых сигналов с частотой дискретизации порядка 12–16 КГц.

Библиографический список

1. Härmä, A. A comparison between frequency-warped and conventional linear predictive coding / A. Härmä, U.K. Laine. – IEEE Trans. Speech and Audio Processing, July 2001.
2. 16 kb/s wideband CELP-based speech coder using mel-generalized cepstral analysis / K. Koishida, G. Hirabayashi, K. Tokuda, T. Kobayashi. – IEICE Trans. Inform. and Systems. – April 2000. – Vol. E83-D, № 4. – P. 876–883.
3. Основы цифровой обработки сигналов: курс лекций / А.И. Солонина, Д.А. Улахович, С.М. Арбузов, Е.Б. Соловьева. – СПб.: БХВ-Петербург, 2005. – 768 с.

[К содержанию](#)