

МЕТОД ЦИФРОВОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ ПРИ ДЕКОДИРОВАНИИ В СИСТЕМАХ СВЯЗИ С НЕЛИНЕЙНОЙ ИКМ

Кузнецов М.В.

Поволжская государственная академия телекоммуникаций и информатики
Самара

Реферат Предложен метод повышения помехоустойчивости приёма речевых сигналов в системах связи с нелинейной ИКМ в каналах с большим уровнем помех, когда появляются импульсные помехи, воспринимаемых на слух как «щелчки». Метод обнаружения щелчков основан на сравнении текущего уровня первой производной речевого сигнала с адаптивным порогом. При исследовании эффективности предложенного метода было проведено обширное моделирование на компьютере. После обнаружения и коррекции аномальных цифровых ошибок предложенным методом вероятность остаточных ошибок не превышает 10^{-5} , а на слух «щелчки» исчезли полностью, что позволяет использовать канал, не удовлетворяющий нормам, для цифровой передачи речи.

Введение

В настоящее время для передачи цифровой речевой информации широко используется нелинейная импульсно-кодовая модуляция (логарифмическая ИКМ). Искажения знакового и старших разрядов номера сегмента нелинейного ИКМ - кода (аномальные цифровые ошибки) вызывают характерные импульсные помехи, воспринимаемые на слух в виде «щелчков». Допустимая вероятность ошибки составляет не более одного «щелчка» в минуту, $P_{\text{ош.доп.}} < 10^{-6}$ [1]. Снижение отношения сигнал/шум в два раза (на 3 дБ) приводит к увеличению $P_{\text{ош}}$ примерно на 3 порядка, или вероятности ошибки $P_{\text{ош.}} = 10^{-3}$, при этом происходит 16 щелчков в секунду, что совершенно недопустимо. Благодаря обнаружению и коррекции этих импульсных помех можно достичь субъективного качества восприятия речи, соответствующего вероятности ошибки $P_{\text{ош.}} = 10^{-5} \div 10^{-6}$. Следовательно, коррекция щелчков снижает требования к отношению сигнал/шум на 2÷3 дБ, что позволяет увеличить допустимую длину участка регенерации, либо повысить относительное количество пар кабеля, уплотняемых системами ИКМ, либо использовать более дешёвый кабель.

Предмет исследования

Предложенный метод обнаружения щелчков основан на сравнении текущего цифрового уровня первой производной речи с двумя адаптивными цифровыми порогом разного знака. Для нахождения производной берётся разность кодов двух соседних отсчётов линейной ИКМ. Величина порогов пропорциональна коду среднего модуля первой производной, накопленному на предыдущем интервале анализа.

Для коррекции «щелчков» были исследованы два вида интерполяции: нулевого порядка, т.е. код обнаруженного пораженного отсчета заменяется кодом предыдущего, и первого порядка, когда код пораженного отсчета берётся как полусумма двух соседних с ним. Чтобы оценить эффективность видов интерполяции, они сравнивались с идеальной интерполяцией бесконечного порядка, когда код пораженного отсчета заменяется истинным.

Аномальные цифровые ошибки являются одиночными, поскольку в цифровом тракте происходит перемежение символов группового сигнала и пачка ошибок в групповом тракте приводит к одиночным ошибкам в индивидуальных каналах ТЧ. Метод обнаружения и коррекции «щелчков» может быть реализован в многоканальном варианте на входе группового ЦАП, до индивидуальных ФНЧ.

Для исследования эффективности предложенного метода было проведено обширное моделирование на компьютере. На примере пяти мужских и трёх женских голосов анализировалось влияние порядка интерполяции, величины порогов обнаружения ошибки, длительности интервала анализа, вида сигнала (речь или пауза), а так же вероятности ошибочного приёма символа в кодовой комбинации.

В отрезки речи длиной 1 минуту (около 500 тыс. отсчётов, т.е. $4 \cdot 10^6$ бит) вносились одиночные случайные ошибки с вероятностью $P_{\text{ош}} = 10^{-2} \div 10^{-4}$. Интервал анализа изменялся от 4 мсек (32 отсчёта) до 16 мсек (128 отсчётов). Величины порогов обнаружения изменялись в диапазоне от 1 до 6 средних модулей первой производной сигнала. Оптимизировалась величина порогов в зависимости от вида сигнала (мужской голос, женский голос, паузы речи).

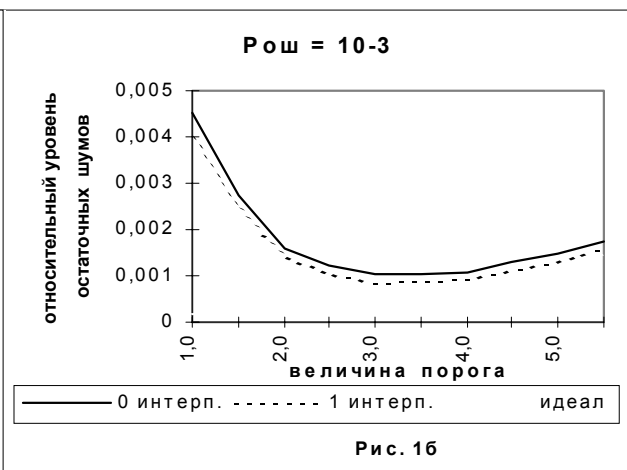
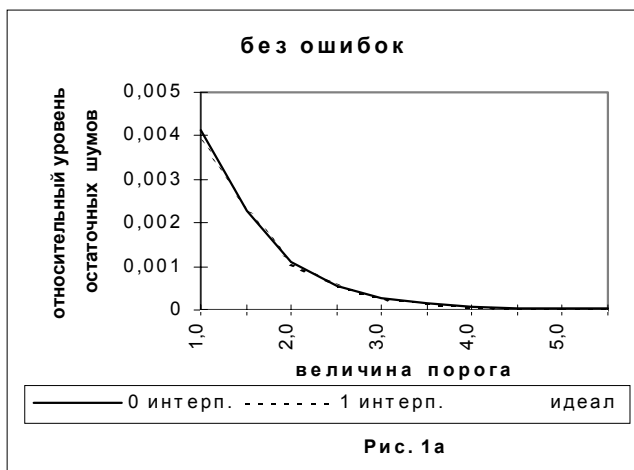
Эффективность обнаружения и коррекции «щелчков» оценивалась по относительному уровню остаточных шумов δ .

$$\delta = \frac{P_{\text{щелчки}}}{P_{\text{сигнал}}}, \quad m_{eX} = \sum_{i=1}^N (x_i)^2, \quad m_{\text{испр}} = \sum_{i=1}^N (x_i - x_{i(i-p)})^2, \quad \text{где}$$

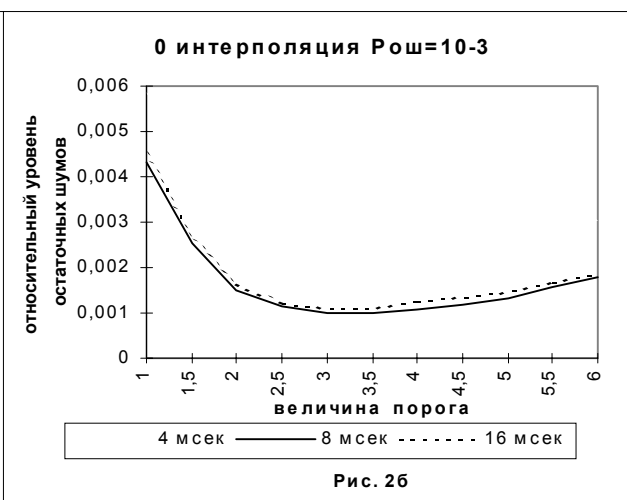
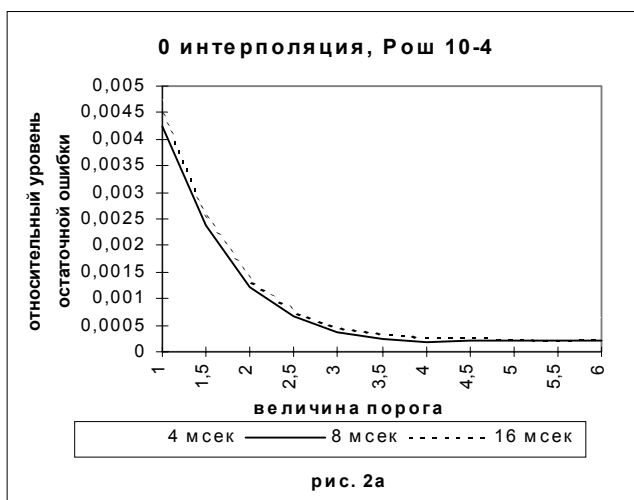
x_i - истинное значение i -го отсчёта сигнала,

$x_{i(\text{испр})}$ - величина откорректированного отсчёта,

N - общее количество отсчётов.

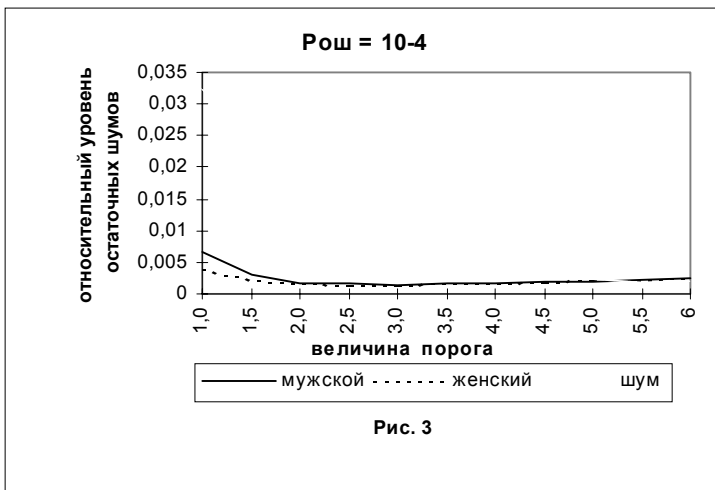


На рис. 1 представлены диаграммы относительного уровня остаточных шумов для интерполяции нулевого, первого и бесконечного порядка. Речевой сигнал имеет большую избыточность и разность между соседними отсчётами невелика [2]. Относительная разница между использованием интерполяции нулевого и первого порядка незначительна и не превышает (1÷1,5)дБ, а между интерполяциями первого и бесконечно высокого порядков не превышает 1дБ. Схемотехническая реализация интерполяции первого порядка намного сложнее, чем нулевого, следовательно предпочтительнее использовать нулевой порядок интерполяции как наиболее оптимальный.



На рис. 2 показана зависимость относительного уровня остаточных шумов от длины интервала анализа. Видно, что при коротком интервале в 4 мсек не происходит достаточного усреднения, в следствии этого средние модули производной двух соседних интервалов сильно отличаются и порог получается недостоверным, а при длинном интервале в 16 мсек проявляется нестационарность речи и информация о пороге, полученная на предыдущем интервале, устаревает. Интервал анализа длиной 8 мсек является оптимальным.

Наличие минимума графиков рис.1-3 объясняется тем, что при низких порогах ошибки хорошо отслеживаются, но так же происходит ложное исправление самого сигнала (вносятся частотные искажения - ослабляются верхние частоты), а по мере роста величины порога всё большая часть ошибок оказывается ниже порога и не исправляется. То есть должен быть оптимальный порог, при котором количество ложных исправлений и необнаруженных ошибок минимально.



После разделения сигнала по голосам и отделения речи от пауз стало ясно, что оптимальные адаптивные пороги для мужских и женских голосов практически одинаковые и от параметров голоса не зависят. Для шума в паузах речи целесообразно принять более высокий порог величиной 5 средних модулей шума не зависимо от уровня сигнала, как это видно из рис. 3.

Полагая закон распределения мгновенных значений первой производной речевого сигнала гауссовым, целесообразно установить адаптивные пороги на уровне порядка $\pm 3U_{эфф}$, где $U_{эфф}$ - эффективное (действующее) значение первой производной речевого сигнала, измеренное на предыдущем отрезке сигнала. Поскольку измерение действующего значения случайного процесса связано со значительными трудностями, гораздо проще измерить средний модуль сигнала $U_{ср}$, пропорциональный действующему значению. Для гауссова процесса известна связь

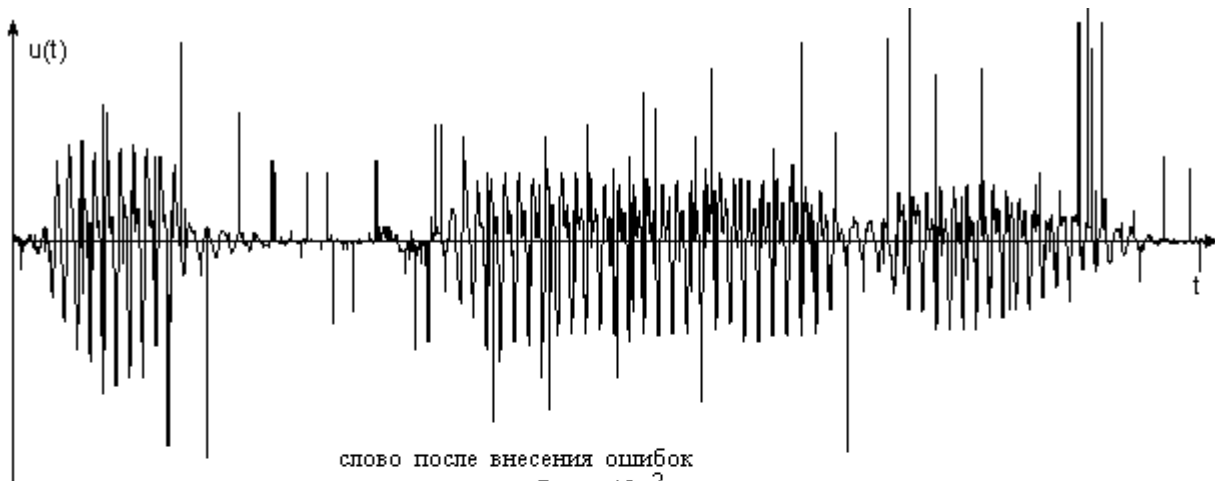
$$U_{ср} = \sqrt{2/\pi} \cdot U_{эфф} \approx 0,8 U_{эфф}$$

Отсюда оптимальная величина адаптивного порога для речи, и это подтверждается результатами обширного моделирования на ПК с использованием реального речевого сигнала и усреднением по большому числу мужских и женских голосов:

$$U_{опт} = 3 U_{эфф} \approx 3,75 U_{ср}$$

Как видно из вышеприведённых диаграмм, реальные значения порогов соответствующие минимуму остаточных шумов, близки к теоретическому $U_{опт}$.

На рисунке 4а представлен фрагмент речевого сигнала (слово «четыре» длительностью 0,616 сек, примерно 40 кбит), в который были внесены ошибки с вероятностью $P_{ош}=10^{-2}$, это примерно соответствует 160 «щелчкам» в секунду (что вообще недопустимо для работы цифрового канала). После обработки этого сигнала предложенным методом вероятность остаточной ошибки составила примерно $P_{ош}=10^{-5}$, а на слух полностью исчезли все «щелчки» при достаточно хорошем качестве звучания речи. Исправленный сигнал представлен на рисунке 4б.



слово после внесения ошибок
с вероятностью $P_{\text{ош}} = 10^{-2}$

Рис. 4 а



слово после исправления ошибок
(интерполяция нулевого порядка)

Рис. 4 б

Выводы

После обнаружения и коррекции аномальных цифровых ошибок предложенным методом можно достичь субъективного качества восприятия речи, соответствующего вероятности ошибки $P_{\text{ош}} = 10^{-5} \div 10^{-6}$. На слух «щелчки» исчезают полностью, что позволяет использовать канал, не удовлетворяющий нормам с $P_{\text{ош}} = 10^{-2} \div 10^{-5}$, для цифровой передачи речи.

Интерполяция нулевого порядка является достаточно эффективной, проигрыш по уровню остаточных шумов по сравнению с интерполяцией первого порядка не превышает $1 \div 1,5$ дБ, а сравнительно с интерполяцией бесконечного порядка $< 2,5$ дБ.

Длительность интервала анализа 8 мсек (или 64 отсчёта) оказалась оптимальной, поскольку более короткий интервал не позволяет достичь надёжного усреднения и адаптивный порог получается недостоверным. При более длинном интервале анализа проявляется нестационарность речевого сигнала и информация о среднем модуле первой производной, полученная на предшествующем интервале, уже устаревает.

Для речи адаптивные пороги составляют $3,5 \div 4$ среднего модуля первой производной, что подтверждает предположение о нормальном распределении мгновенных значений первой производной речевого сигнала. Порог в паузах речи целесообразно принять равным 5 средним модулям шума.

Список литературы

1. Гуревич В.Э., Лопушнян Ю.Г., Рабинович Г.В. Импульсно-кодовая модуляция в многоканальной телефонной связи. - М.: Связь, 1973.
2. Ротенштейн И.В. Исследование вокализованной речи и двух ее первых производных на интервале локальной стационарности. Сборник трудов ученых Поволжья (материалы НТК ПГАТИ). Выпуск № 4, Самара, 1999.



THE METHOD OF DIGITAL PROCESSING OF SPEECH SIGNALS WHEN DECODING IN COMMUNICATIONS NETWORKS WITH LOGARITHMIC PCM

Kuznetsov M.V.

state academy of telecommunications and informatics
Russia, Samara

Abstract. Offered method of raising noise-immunity a receiving the speech signals in communications networks with nonlinear PCM in channels with the big hindrance level, when appear pulsed hindrances, perceived on the rumour as «snub». Method of finding the snubs is based on the comparison of current first derived speech signal level with the adaptive threshold. At the efficiency study of offered method was conducted extensive modeling on the computer. After finding and correcting the anomalous numerical mistakes probability of offered by the method remaining mistakes does not exceed 10^{-5} , but on the rumour «snub» disappeared completely that allows to use a channel, fail the standards, for the numerical issue of speech.

Introduction

Garbling the senior bits of segment number nonlinear PCI - a code cause distinctive pulsed hindrances, perceived on the rumour in the manner of «snubs». Possible probability of error forms not more one «snub» at a minute, $D_{er} < 10^{-6}$. Due to finding and correction of these pulsed hindrances possible to reach a subjective quality of perception of speech, corresponding probability of error $D_{er} = 10^{-5} \div 10^{-6}$. Consequently, correcting the snubs reduces requirements to the attitude a signal/noise on 2÷3 dB.

Subject of research

Offered method of finding the snubs is based on the comparison of current first derived speech level with two adaptive thresholds of different sign. Value of thresholds proportional average modula first derived, accumulated on the preceding interval of analysis.

For the efficiency study of offered method was conducted extensive modeling on the computer. On the example of seven men and three women voices was analysed influence of order to interpolations, values of thresholds of finding an error, duration of interval of analysis, type of signal (speech or pause), and the same way probability wrong receiving a symbol in code combinations. In the length of speech by the length 1 minute (near 500 thousand code combinations , i.e. $4 \cdot 10^6$ bit) were contributed single casual errors with probability $D_{er} = 10^{-2} \div 10^{-4}$. Interval of analysis was changed from 4 ms (32 code combinations) before 16 ms (128 code combinations). Values of thresholds of finding were changed within the range of from 1 before 6 average modulas of first derived signal. Optimized value of thresholds depending on the type of signal (male voice, feminine voice, pauses of speech).

Efficiency of finding and correction «snubs» was valued on the relative remaining noise level δ .

$$\delta = \frac{P_{er.OUT}}{P_{IN}}, \quad P_{IN} = \sum_{i=1}^N (x_i)^2, \quad P_{er.OUT} = \sum_{i=1}^N (x_i - x_{i(COR)})^2, \quad \text{where}$$

x_i - a true value i-s code combination a signal,

$x_{i(COR)}$ - a value corrected code combination ,

N - amount code combinations.

On the Fig 1a presented fragment of speech signal (word «четыре» duration 0,616 sek, approximately 40 kbit), in which were contributed errors with probability $D_{er} = 10^{-2}$, this approximately corresponds 160 «snubs» at a second (that in general inadmissible for functioning (working) a digital channel). After processing this signal probability of offered by the method remaining error has formed approximately $D_{er} = 10^{-5}$, but on the rumour completely disappeared all «snub» under sufficiently good a quality of sounding a speech. Corrected signal is submitted for the Fig 1b.

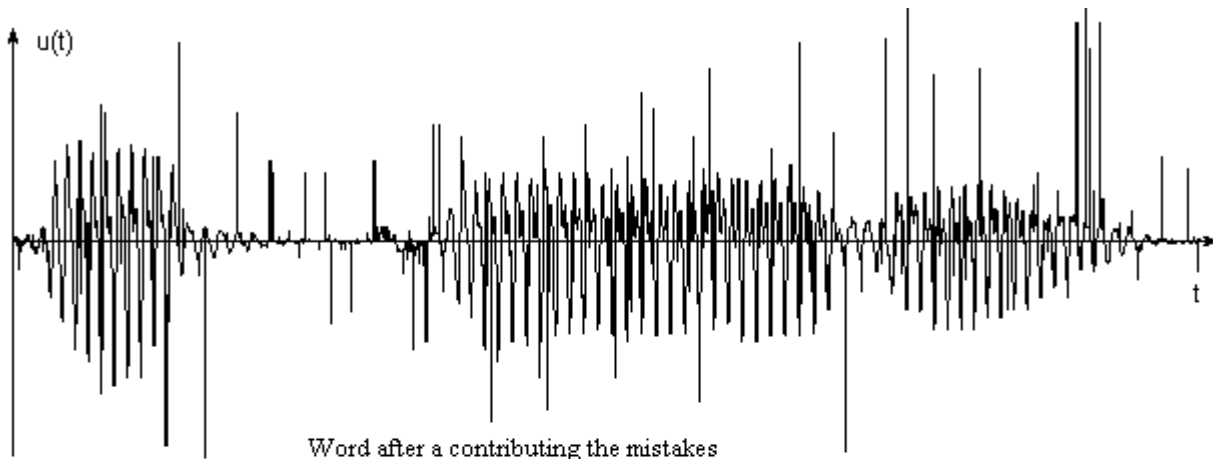
Conclusions

After finding and correcting the anomalous numerical mistakes offered by the method possible to reach a subjective quality of perception of speech, corresponding probability of mistake $D_{er} = 10^{-5} \div 10^{-6}$. On the rumour «snub» disappear completely that allows to use a channel, fail the standards with $D_{er} = 10^{-2} \div 10^{-5}$, for the numerical issue of speech.

Interpolation of zero order is sufficiently efficient, loss on the remaining noise level in contrast with the first-order interpolation does not exceed 1÷1,5 dB, but relatively with the interpolation of endless order < 2,5 dB.

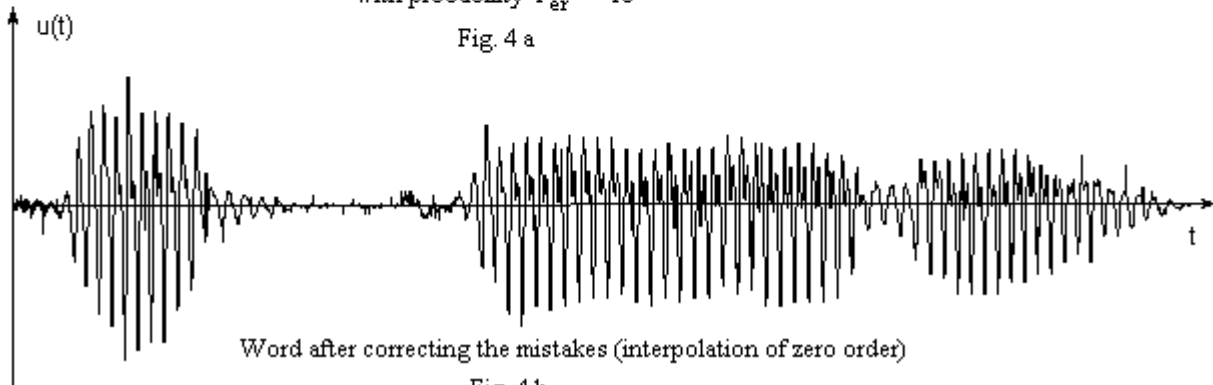
Duration of interval of analysis 8 ññ (or 64 counting out) is rendered optimum, as far as more short interval does not allow to reach a reliable averaging and adaptive threshold is got unauthentic. Under the longer interval of analysis reveals itself variability speech signal and information on the average modula first derived, tinned on preceding interval, already grows old.

For the speech adaptive thresholds form 3,5÷4 average modula first derived that confirms a suggestion on normal sharing the instant values of first derived speech signal. Threshold in pauses of speech reasonable to take equal 5 average modulas of noise.



Word after a contributing the mistakes
with probability $P_{er} = 10^{-2}$

Fig. 4 a



Word after correcting the mistakes (interpolation of zero order)

Fig. 4 b

References

1. Гуревич В.Э., Лопушнян Ю.Г., Рабинович Г.В. Импульсно-кодовая модуляция в многоканальной телефонной связи. - М.: Связь, 1973.
2. Ротенштейн И.В. Исследование вокализованной речи и двух ее первых производных на интервале локальной стационарности. Сборник трудов ученых Поволжья (материалы НТК ПГАТИ). Выпуск № 4, Самара, 1999.