Заключение

Результаты проведенных исследований позволяют сделать следующие основные выводы:

1. Использование медианных фильтров позволяет восстанавливать речевые сигналы, искаженные импульсным шумом с фиксированными и случайными значениями импульсов.

2. Взвешенный медианный фильтр является наиболее эффективным алгоритмом удаления импульсного шума среди рассматриваемого класса медианных фильтров.

3. Найдены оптимальные параметры взвешенного медианного фильтра с точки зрения критериев PESQ и MSE для двух речевых эталонов с разным динамическим диапазоном. В двух случаях наблюдается расхождение этих параметров.

4. В случае импульсного шума с фиксированными значениями импульсов имело место расхождение в оценке качества сигнала по критериям PESQ и MSE. По критерию PESQ при плотности шума более 7,7% фильтрация переставала оказывать улучшающее воздействие, в отличие от оценки по критерию MSE.

5. При удалении импульсного шума из речевых сигналов следует руководствоваться параметрами фильтров, оптимизированными по критерию PESQ, как наиболее точному инструменту эталонной оценки качества речи на данный момент развития телекоммуникационных систем.

Литература

1. Brownrigg D. The weighted median filter // Comm. ACM. № 27, P. 807-818, 1984.
2. Радзишевский А.Ю. Основы аналогового и цифрового звука – М.: Вильямс, 2006.
3. Benesty J., Sondhi M., Huang Y. Handbook of Speech Processing, Springer-Verlag, 2008.
4. Морриси П. Как измерить качество речевой связи // Сети и системы связи, № 8, 2005.
5. Rix A., Hollier M., Hekstra A., Beerends J.: Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment, Part 1 // 2002 (www.psytechnics.com).
6. Rix A., Hollier M., Hekstra A., Beerends J.: Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment, Part 2 // 2002 (www.psytechnics.com).

IMPULSE NOISE REMOVING FROM SPEECH SIGNALS USING WEIGHTED MEDIAN FILTERS

Gerasimov N., Kuykin D., Khryashchev V.

Yaroslavl Demidov State University

Often median filters are used for digital signals preliminary processing and restoration because of it's ability to decrease outliers while staying original samples non modified. It makes possible to apply median filters for impulsive noise removal from images and speech signals [1].

Using of weighted median filters provides more possibilities of filter parameters attenuation for specified type of noise removing. Comparative analysis of median and weighted median filters for speech signals denoising is presented using signal quality criteria PESQ [2]. Salt-and-pepper and random valued impulsive noise are considered.

Received results of salt-and-pepper impulsive noise corrupted test signal "Car" restoring by different algorithms are depicted in Table 1.

Table 1.

Filter type	PESQ	MSE*10^{-4}
Corrupted by 2% salt-and pepper noise signal	1.745	197
Median filter (3*1)	2.192	9.07
Median filter (5*1)	**2.942**	**3.78**
Median filter (7*1)	2.743	6.6
Median filter (9*1)	2.464	11.2
Weighted median filter (121)	1.975	55.7
Weighted median filter (13531)	2.448	6.72
Weighted median filter (1357531)	**3.010**	**3.26**
Weighted median filter (135797531)	3.005	3.47

Random valued impulsive noise removing from speech signals were considered also and similar results were obtained.

As results of researches optimised weighted median filter parameters for effective impulse noise removal were obtained for two types of impulsive noise. The optimised mask size for two cases of impulse noise elimination are presented in Table 2.

Table 2.

Noise type	Test signal	Filter mask size	
		PESQ	MSE
Salt-and-pepper impulsive noise	"Car"	7*1	7*1
	"ENG_M"	7*1	5*1
Random valued impulsive noise	"Car"	7*1	5*1
	"ENG_M"	5*1	5*1

As a result of our work recommendations for choice of mask size and weights of weighted median filters depending on noise level was generated based on PESQ signal quality criteria.

References

1. Brownrigg D. The weighted median filter // Comm. ACM. № 27, P. 807-818, 1984.
2. Rix A., Hollier M., Hekstra A., Beerends J.: Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment, Part 1 // 2002 (www.psytechnics.com).

ОБЪЕКТИВНЫЕ ОСНОВЫ ПОВЫШЕНИЯ ЕСТЕСТВЕННОСТИ (НАТУРАЛЬНОСТИ) СИНТЕЗИРОВАННОЙ РЕЧИ ПРИ РАСШИРЕНИИ ПОЛОСЫ ЧАСТОТ РЕЧЕВОГО СИГНАЛА ДО ДИАПАЗОНА 50 – 7000 ГЦ

Рыболовлев, А.А., Илюшин М.В.

Академия Федеральной Службы Охраны Российской Федерации, г. Орел

В настоящее время все мы являемся свидетелями и даже участниками процесса построения глобального информационного общества, основой функционирования которого является возможность предоставления пользователям широкого спектра современных инфокоммуникационных услуг в любое время при нахождении абонентов на стационарных объектах и в движении. Данная возможность обеспечивается за счет конвергенции сетей связи общего пользования и таких технологий, как сети сотовой связи поколения 3G и Интернет. Дальнейшее развитие в направлении объединения предоставляемых услуг привело к формированию концепции сетей связи следующего поколения – NGN (Next Generation Network).

Объем передаваемой в мире информации постоянно растет. Согласно имеющимся в научной литературе данным, период удвоения циркулирующего по сетям объема информации в мире сокращается: с 5 лет в 1980 г. до 3 мес. в настоящее время, причем ожидается ускорение этого процесса. Анализ прогнозов роста объема доходов и количества подписчиков по основным услугам сетей мобильной связи поколения 3G позволяет оперировать следующими цифрами. Прогнозируемый среднегодовой темп роста (СГТР) количества пользователей телефонными услугами сетей связи 3G в период с 2005 г. по 2010 г. составляет 46%, а СГТР объема доходов от традиционных услуг по передачи речи за тот же период равен 32%. Также прослеживается тенденция увеличения телефонных услуг с улучшенным качеством. Ожидаемый СГТР объема доходов от передачи высококачественной речи в период с 2005 г. по 2010 г. составляет 95% [2,3,6].

Необходимо отметить, что существующие и перспективные технологии в рамках развития сетей связи следующего поколения, как правило, не ориентированы на использование низкоскоростных каналов из-за сравнительно невысокого качества восстановленного речевого сигнала (РС). Дополнительные проблемы при построении низкоскоростных систем передачи возникают в тех случаях, когда в системе связи требуется обеспечить конфиденциальность передаваемой информации, а также устойчивость работы системы в целом при изменениях ее структуры и параметров, т.е. при функционировании ведомственных систем связи для нужд государственного управления.

Качество телефонных услуг, предоставляемых абоненту, в основном определяется алгоритмом кодирования в речепреобразующем устройстве. Целью кодирования речи является получение компактного цифрового описания РС в форме, которая может быть использована для эффективной записи и передачи его в виде кодированного (сжатого) сигнала.

Исторически сложились три подхода к технологии преобразования речи.

В *кодерах формы сигнала* кодируется форма РС как функция времени и при достаточно высокой скорости передачи обеспечивается высокое качество восстановленной речи. При *параметрическом кодировании* моделируется процесс речеобразования человека. Для этого в кодере из речевого сигнала вычисляются определенные параметры, которые передаются к декодеру, где они используются для восстановления формы сигнала. Использование полностью параметрических методов в настоящее время ограничено, так как они приводят к заметному ухудшению натуральности звучания голоса. Один из способов снижения скорости передачи речи и повышения эффективности использования полосы пропускания канала связи состоит в применении *гибридных методов*, основанных на принципах линейного предсказания и объединяющих параметрическое кодирование и аппроксимацию формы речевой волны [1].

Выбор кодера для конкретных применений зависит от учета ряда характеристик, к которым относятся:

- вид электросвязи;
- скорость передачи;
- приемлемый уровень качества;
- ограничения на временные задержки;
- учет потерь в канале;
- учет возможности последовательного соединения кодеков при взаимодействии с другими системами передачи.

Речевые сигналы являются случайными, и их особенности выражаются некоторыми видами характеристик. Фонетические характеристики определяют звуковой состав речи. Информационные

характеристики позволяют разделить речевую информацию на сигнальную (определяет источник звука), семантическую (передает содержание речи) и эстетическую (отображает эмоциональные переживания диктора). Временные характеристики определяют длительность различных звуков речи и пауз. К акустическим характеристикам относятся такие физические параметры РС, как его мощность, динамический диапазон, формантный состав, направленные свойства и др.

Использование электрического тракта для передачи речевых сигналов часто приводит к заметному изменению их акустических характеристик. Это не только снижает общее качество звучания, но и сказывается на фонетических характеристиках речи. Трансформация акустических параметров сигнала влияет и на информативные показатели речи, делая ее недостаточно разборчивой и мало выразительной. Все это заставляет более детально исследовать акустические характеристики РС, изменения которых определяют конечное качество восстановленной речи по признаку естественности (узнаваемости, натуральности).

Измерения показывают [5], что звуки речи значительно отличаются по мощности. Так, для гласных звуков средняя мощность составляет 700 мкВт, тогда как для согласных она приближается к 0,7 мкВт. Такое большое различие в мощностях гласных и согласных приводит к снижению разборчивости речи.

Важными факторами ощущаемого качества кодера по признаку естественности восстановленной речи являются частотный диапазон, в котором передается кодируемый сигнал, и формантный состав речи. Решающими в выборе полосы 0,3 – 3,4 кГц были экономические соображения и нехватка телефонных каналов.

Большая часть энергии чаще всего содержится в гласных звуках, которые занимают полосу частот ниже 3 кГц. Однако, всем известно, что согласные звуки являются более информативными по сравнению с гласными. Например, в слове «посылка» звуки *«п», «с», «л», «к»* дают большее представление о его смысле, чем звуки *«о», «ы», «а»*. А для передачи согласных звуков часто требуется учитывать полосу частот выше 3 кГц. Поэтому применение узкополосных систем связано с ухудшением разборчивости, например, звуки *«с» и «ф»* различаются только из-за формант, расположенных в верхней полосе частот.

Согласно результатов исследований в области обработки речевых сигналов [1,5,7] формантные области большинства звуков русского языка находятся в пределах от 100 до 8000 Гц. При этом основные форманты, определяющие распознавание каждого звука, концентрируются в пределах 200-3200 Гц, а вспомогательные форманты, отвечающие за натуральность звучания восстановленной речи, занимают полосу частот от 3000 до 8000 Гц.

Из всего вышесказанного можно сделать вывод, что кодирование узкополосной речи связано со следующими недостатками:

- различие в мощностях гласных и согласных звуков (около 30 дБ) приводит к снижению разборчивости речи;
- у большинства согласных звуков, обладающих большей информативностью по сравнению с гласными, усиленные участки спектра расположены в полосе частот выше 3 кГц;
- вспомогательные форманты остаются за границами используемой полосы частот;
- ухудшение разборчивости по причине ограничения частотного диапазона приводит к увеличению концентрации внимания абонента и, следовательно, к повышению усталости.

В качестве одного из возможных направлений устранения данных недостатков и повышения качества телефонной связи в сетях ведомственного предназначения, необходимость которого обусловлена возрастающими современными требованиями абонентов, может рассматриваться переход от кодирования узкополосного РС (УРС) к передаче широкополосного речевого сигнала (ШРС) с диапазоном частот от 50 до 7000 Гц [1,4].

Разработки в области преобразования ШРС нашли отражение в рекомендациях МСЭ G.722, G.722.1, G.722.2 [8,9,10]. В [4] предлагается модель широкополосного CELP-кодера с реконфигурацией структуры кодовой книги на основании информации, полученной из психоакустической модели.

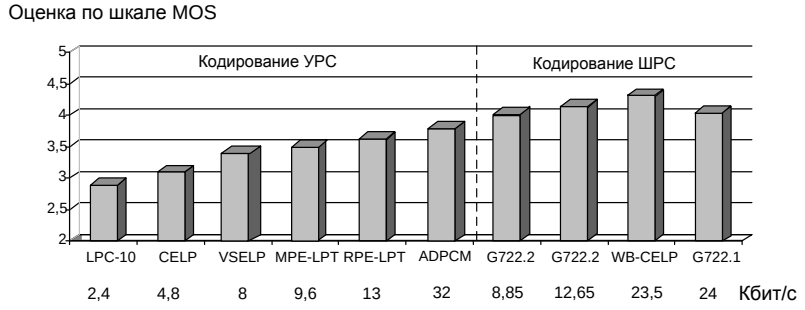

Рис.1. Диаграмма качества речи по пятибалльной шкале MOS, достигаемого различными гибридными кодеками

Из рис.1, на котором показана сравнительная оценка качества речи, обеспечиваемого гибридными кодеками на скоростях до 32 кбит/с, видно, что алгоритмы кодирования ШРС позволяют достичь лучшего качества по сравнению с кодеками УРС на приблизительно равных и даже меньших скоростях.

Перспективным направлением дальнейших исследований в области передачи широкополосного речевого сигнала следует назвать повышение степени адаптивности кодеков к параметрам анализируемого кадра речевого сигнала. Сокращение статистической избыточности речи планируется достичь за счет структурной адаптации системы кодирования к статистическим характеристикам кодируемых параметров сегментов речевого сигнала, разделенным на конечное число классов. Введение в состав кодера психоакустической модели и трехмерной кодовой книги, отражающей свойства кодируемого сигнала, позволит исключить перцептуальную избыточность, снизив тем самым требования к объему передаваемой информации. При разработке кодека широкополосного РС должны быть учтены аспекты, связанные с учетом характеристик ШРС, возможностью субполосного липредерного кодирования и способностью существующих цифровых процессоров обеспечить функционирование перспективного кодека речевых сигналов в режиме реального времени [1,4].

Таким образом, на основе анализа направлений развития в области кодирования речевых сигналов есть основание считать, что в настоящее время существуют все предпосылки для перехода к передаче широкополосного речевого сигнала с диапазоном от 50 до 7000 Гц. Данное расширение полосы частот позволит значительно повысить качество восстановленного речевого сигнала по признаку естественности звучания при ориентации на низкоскоростные каналы передачи в ведомственных сетях.

Литература

1. Быков С.Ф., Журавлев В.И., Шалимов И.А. Цифровая телефония: Учеб. пособие для вузов. – М.: Радио и связь, 2003. – 144 с.: ил.

2. Громаков Ю. А. Концептуальные аспекты развития сотовой связи // Электросвязь. – 2003. – № 11. – С. 65 – 70.

3. Гулевич, Д.С. Сети связи следующего поколения: учеб. пособие для вузов / Д.С. Гулевич. - М.: Интернет-университет Информационных Технологий: БИНОМ, 2007. - 183 с. : ил., табл.

4. Лившиц, М. З. Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой / М. З. Лившиц, М. Парфенюк, А. А. Петровский // Цифровая обработка сигналов. – 2005. – № 2. – С. 20 – 35.

5. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. – М.: Радио и связь, 1987. – 168 с.

6. Москвитин В.Д. Рост объемов информации – главный фактор развития пакетных сетей // Электросвязь. – 2008. – № 10. – С. 32 – 33.

7. Попов О.Б., Рихтер С.Г. Цифровая обработка сигналов в трактах звукового вещания : Учебное пособие для вузов. – М.: Горячая линия – Телеком, 2007. – 341 с.

8. ITU-T Recommendation G.722. 7 kHz audio-coding within 64 kbit/s. – Geneva, 1988.

9. ITU-T Recommendation G.722.1. Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss. – Geneva, 1999.

10. ITU-T Recommendation G.722.2. Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB). – Geneva, 2003.

OBJECTIVE BASES OF INCREASE OF NATURALNESS (RECOGNITION) OF THE SYNTHESIZED SPEECH AT EXPANSION OF THE STRIP OF FREQUENCIES OF THE SPEECH SIGNAL TO THE RANGE 50 - 7000 HZ

Rybolovlev A., Iljushin M.

Academy of FGS of Russia

Now in the field of telecommunications there is a process of transition to networks which give users a wide spectrum of qualitative services. It is necessary to note, that existing and perspective technologies of batch transfer of the information are not focused on use of channels by low speed of transfer owing to rather low quality of the synthesized speech signal.

In this article the lacks arising at coding of a standard speech signal with a strip of frequencies 0,3 - 3,4 kHz are described. To them concern:

- distinction in capacities of vowels and consonants results in decrease in legibility of speech;
- at the majority of the consonants possessing the greater information, than the vowels, amplified sites of a spectrum are located in a strip of frequencies above 3 kHz;
- auxiliary amplified sites remain behind borders of a used strip of frequencies;
- deterioration of legibility owing to restriction of a frequency range results in increase in concentration of attention of the subscriber and, hence, to increase of weariness.

As one of possible directions of elimination of the given lacks and improvement of quality of telecommunication in networks of departmental applicability which necessity is caused by growing modern requirements of subscribers, transition from coding narrow-band speech to transfer of a wideband speech signal with a range of frequencies from 50 up to 7000 Hz can be considered.

The analysis of transformation of speech with the help of hybrid algorithms allows to draw a conclusion, that codecs of a wideband speech signal provide better quality in comparison with systems of compression of narrow-band speech on approximately equal and even smaller speeds. Expansion of a strip of frequencies will allow to increase considerably quality on the basis of naturalness of sounding of a restored speech signal.

АНАЛИЗ И ВЫБОР ЧАСТОТЫ ДИСКРЕТИЗАЦИИ ДЛЯ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

Фатуллаев А.Б., Ибрагимов Б.Г.
Институт Кибернетики НАНА, Баку
Азербайджанский Технический Университет, Баку

В современном этапе развития систем передачи и обработки непрерывных сигналов при распознавании речевой информации на базе перспективных лингвистических DSP (Digital Signal Processing) технологий приобретают большую актуальность в системах цифровым методом обработки речевых сигналов. При этом большой интерес вызывает непрерывных речевых сигналов по некоторым входным данным, таких как амплитудно-частотных и амплитудно-фазовых характеристик распознаваемых речевых сигналов посредством преобразования аналогового сигнала в цифровой сигнал.

Известно [1], что цифровое представление непрерывных сигналов обеспечивает эффективную предварительную обработку речевых сигналов, помехоустойчивость и надежность связи, а также возможность защиты от несанкционированного доступа путем засекречивания.

В системах обработки и распознавания речи используются устройства ввода и передачи непрерывных сигналов, которые реализуют предварительную обработку входной речевой информации с целью получения более компактного описания входного речевого сигнала.

Однако, проведенные экспериментальные исследования показали, что ранее полученная частота дискретизации, шаг квантования и длина двоичной кодовой комбинации не всегда удовлетворяют вышеизложенным требованиям по предварительной обработке речевых сигналов в частотной области для распознавании речи.

На повышения эффективности распознавания речевой информации посвящены разные труды зарубежных специалистов многих стран [2,3], начиная от ввода, предварительной обработки и до озвучивания речи. Их целью является создание систем речевого общения между человеком и компьютером, а также методы анализа и синтеза алгоритмов работы систем распознавания речевых сообщений.

На основе системно-технического анализа установлено [4,5], что один из существенных факторов влияющих на ухудшение качества ввода и передачи речи является ее этапная цифровая обработка. Несмотря на то, что речь обладает значительной временной избыточностью, ее качество ввода и передачи через устройства ввода цифрового процессора становится в значительной мере зависимой от параметров речевого сообщения.

Учитывая, особенности, состав и характер исследуемого речевого сообщения при преобразовании и обработке непрерывных сигналов можно рассматривать как процесс аппроксимации непрерывных сигналов цифровыми сигналами в широком смысле, т.е. сигналами значениями отсчетов, необходимого для ввода с требуемой достоверностью. Эти системы приобретают важное значение в связи с внедрением в практику низкоскоростных цифровых каналов связи, где скорость передачи $V_k \leq 64$ Кбит/с.

В данной работе рассматривается вопрос выбора эффективного значения частоты дискретизации для речевого сигнала в системах распознавания речи (при обработке непрерывных речевых сигналов) с применением метода квадратурной обработки узкополосных сигналов, т.е. методом субдискретизации сигналов.

Математическая формулировка предложенного подхода для непрерывных речевых сигналов может быть представлена следующей целевой функцией:

$E_д(\Delta F, t_{рс}, L_{дк}) = \{U(t_{рс}), U_d[kT_д], U_k[k\Delta t_{kv}], U[N(km)]\}$, k=0,1,2,… **(1),** где $U(t_{рс})$ – функция амплитуды входного непрерывного речевого сигнала t_c; $U_d[kT_д]$–функция дискретизации входного сигнала по времени в виде дискретного отсчета; $U_k[k\Delta t_{kv}]$ – функция квантования входного дискретизированного сигнала по уровню в виде дискретного отсчета; U[N(km)] – функция двоичного исчисления, учитывающая операцию кодирования квантованных значений передаваемого сигнала в виде последовательности двоичных кодовых комбинаций с длиной $L_{дк}$; m - основании используемого кода.

Выражение (1) является аналитическим алгоритмом преобразования непрерывного сигнала, с помощью которого может быть определено минимально эффективное значение возможной частоты дискретизации $F_д$ преобразуемого речевого сигнала $U(t_{pc})$ по частотным критериям.

Проведенные исследования показали [1,2], что для эффективной передачи (в смысле возможности последующего восстановления без потерь) и распознавания речевого сигнала на базе компьютерных технологий необходима предварительная обработка спектра речевого сигнала, обеспечивающая устойчивость распознавания речи при наличии частотных искажений сигнала из-за этапной цифровой обработки. Поэтому в устройствах цифрового процессора для предварительной обработки речевого сигнала, частотный спектральный анализ речи играет важную роль.

С этой целью, т.е. для предварительной обработки речевого сигнала, на основе проведенного системно-технического анализа, предлагается решение, реализованное на основе структурно-функциональной схемы цифрового процессора (обработки сигналов) с использованием DSP-технологии и состоящая из следующих блоков: анализатор, квантователь и кодер с фильтром VAD (Voice Activate Detection), которая показана на рис.1.

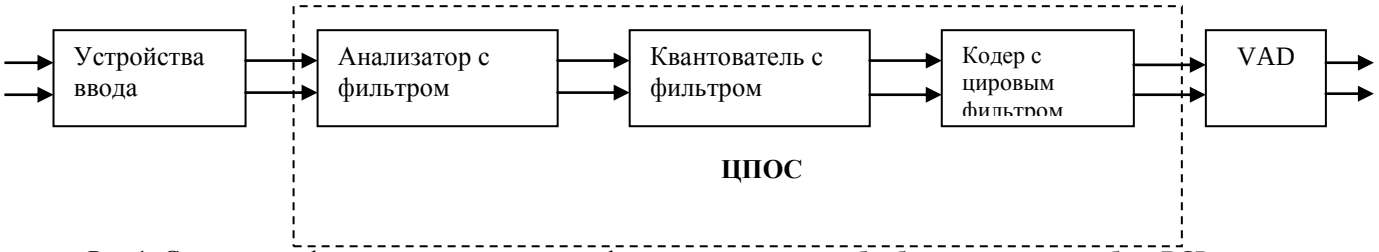

Рис.1. Структурно-функциональная схема цифрового процессора обработки сигналов на базе DSP-технологии

В рассматриваемой структурной схеме важным модулем является цифровой процессор обработки сигналов (ЦПОС). Его составными элементами являются дискретизатор, квантователь и кодер с фильтром, а также VAD. Для реализации алгоритмов преобразования речевой информации, ЦПОС содержит модуль дискретизатор, модуль квантователь и кодер с фильтром. Модуль дискретизатор имеет фильтр низкой частоты, ограничивающий ширину спектров исходного сигнала, который преобразовывает входной аналоговый речевой сигнал $U_{pc}(t)$ в дискретный $U_д(t)$, через интервал дискретизации $\Delta t_д$ в виде $U[kT_д]$, k =0, 1, 2,… дискретных отсчетов. Полученные отсчеты мгновенных значений в модуле квантования квантуются по уровню и превращаются в цифровую кодовую комбинацию на базе двоичного кодера.

Кроме этого, на основе исследований в [2,3] установлено, что для передачи и обработки узко- и широкополосных сигналов важное место занимают методы субдискретизации сигналов, сущность которых составляет метод квадратурной дискретизации сигнала. В телекоммуникационных системах квадратурная обработка сигналов является одним из высокоэффективных методов модуляции непрерывных и цифровых сигналов, которые позволяют дискретизировать полосовой сигнал с частотой, определяемой не верхней границей, а шириной спектра сигнала.

Для решения задачи предварительной обработки частотных характеристик речевых сигналов в распознавании речи, выбран метод квадратурной обработки узкополосных непрерывных сигналов (ΔF=0,3…3,4кГц), т.е. метод субдискретизации сигналов.

Частотный спектральный анализ непрерывных речевых сигналов заключается в следующем:

1. Выбор частоты дискретизации на основе методов квадратурной дискретизации узкополосных сигналов, базирующейся на теореме Котельникова. Для решения данной задачи применен метод субдискретизации сигналов, а алгоритмы реализации выполняются следующим образом:

Пусть аналоговый полосовой сигнал имеет спектр, расположенный в диапазоне между частотами F_{min} и F_{max}. Как правило частота дискретизации должна превышать $2F_{max}$. Однако для точного восстановления сигнала по его дискретным отсчетам необходимо обеспечить отсутствие перекрытия сдвинутых копий спектра. Это дает дополнительную возможность выбора точного значения частоты дискретизации. И восстановление сигнала в данном случае, естественно, должно производиться с помощью полосового фильтра типа VAD.

На основе метода субдискретизации сигналов и алгоритма предложенного подхода, можно определить условия, которые дают возможность дискретизации сигнала таким образом, что при некотором целом k зеркальная половина спектра будет оказаться расположенной между k-й и $(k+1)$ -й сдвинутыми копиями спектра. Отсюда получаем неравенства:

$$- F_{min} + k \cdot F_д < F_1, \qquad - F_{max} + (k+1) F_д > F_{max} \qquad \textbf{(2).}$$

Можно преобразовать эти неравенства в одно двойное неравенство: $2F_{max} - F_д < k \cdot F_д < 2F_{min}$ **(3)**

Отсюда следует, что $2F_{min} > 2F_{max} - F_д$ и, следовательно $F_д > 2 \cdot (F_{max} - F_{min})$

Неравенства (2) и (3) являются необходимыми условиями выбора эффективного значения дискретизации речевого сигнала.

Таким образом, как и в случае квадратурной дискретизации, частота дискретизации ограничена снизу удвоенной шириной спектра сигнала. С учетом этого из (2) можно определить максимально возможное значение k: $$k < \frac{2F_{\min}}{F_Д} < \frac{F_{\min}}{F_{\max} - F_{\min}} \quad \textbf{(4)}$$

Для всех целых k, не превышающих это значение, из двойного неравенства (3) можно определить диапазон возможных значений частоты дискретизации: $$\frac{2F_{\max}}{k+1} < F_Д < \frac{2F_{\min}}{k} \quad \textbf{(5)}$$

В качестве примера рассмотрим дискретизацию сигнала со средней частотой 25 кГц и шириной полосы 4 кГц. Границы занимаемой сигналом полосы частот в этом случае равны $F_{\min} = 22$ кГц и $F_{\max} = 28$ кГц, а минимальное значение k, согласно (3), определяется следующим неравенством: $k < \frac{F_{\min}}{F_{\max} - F_{\min}} = \frac{22}{6} = 3{,}6 \approx 4$

Для целочисленных значений k=1, …,4, удовлетворяющих это неравенство, согласно (4) имеем следующие диапазоны возможных частот дискретизации речевых сигналов [1]:

1. $k = 1: F_Д = \frac{2 \cdot 28}{2} \frac{2 \cdot 22}{1}$ кГц = 28….44,0 кГц
2. $k = 2: F_Д = \frac{2 \cdot 28}{3} .... \frac{2 \cdot 22}{2}$ кГц = 18,66…44,0 кГц
3. $k = 3: F_Д = \frac{2 \cdot 28}{4} .... \frac{2 \cdot 22}{3}$ кГц = 14…14,66 кГц
4. $k = 4: F_Д = \frac{2 \cdot 28}{5} .... \frac{2 \cdot 22}{4}$ кГц = 11,20….11,0 кГц

На основе расчета получены числовые значения частоты дискретизации речевого сигнала. Здесь показаны, что важные спектры дискретизированного сигнала, получается при выборе частот дискретизации, равных 25 кГц ($k = 1$), 18,6 кГц ($k = 2$) , 14,0 кГц ($k = 3$), 11,20 кГц ($k = 4$). При значении k =0, исходя из (4), дает диапазон частоты дискретизации от $2F_{max}$ до бесконечности и, таким образом, соответствует обычной дискретизации сигнала согласно теореме Котельникова.

2. Выбор шага квантования по уровне непрерывных сигналов, обеспечивающие процесс аппроксимации непрерывных сигналов цифровыми сигналами, т.е. сигналами с дискретными значениями отсчетов и определяется следующим неравенством [5]: $\Delta t_{kv} \le T_д$, $T_д = 2\pi / \omega_B$, $\omega_B \in [-\omega_B, +\omega_B]$ **(6),** где ω_B – ширина верхнего частотного спектра квантуемого непрерывного речевого сигнала.

Выполнение условий (5) и (6), на основе теоремы Котельникова позволяет однозначно устранить частотное искажение речевого сигнала при вводе его по дискретным отсчетам $U[kT_д]$, k=0, 1, 2,… включая и гармоническую составляющую сигнала.

3.Выбор кода на основе двоичного счисления для кодирования квантованных сигналов. Запись квантованного уровня с L_{kv} разрешенными уровнями в двоичной системе счисления может быть представлена в виде[1]: $$L_{kv} = \sum_{i=1}^{m} a_{m-i} \cdot 2^{m-i}, \quad i = \overline{1, m} \quad \textbf{(7).}$$

После строгие выполнения выше указанных этапных - частотных, временных и кодированных алгоритмов, позволяют более эффективно реализовать предварительная обработка и ввода цифровых речевых сообщений при распознавании речевых сигналов, которые особое место занимает в системе речевой связи с машиной, т.е. в системе распознавания речи.

Результаты исследования и анализ показали, что полученные могут быть использованы для предварительной обработки и ввода речевых сообщений в системе распознавания речи.

Литераура

1. Сергиенко А.Б. Цифровая обработка сигналов. Учебник для вузов. 2-е изд. – СПб.: Питер, 2007. – 751с.

2. Фатуллаев Ф.Б., Ибрагимов Б.Г. Выборы частоты дискретизации в системах распознавания речи // Труды Международной Академии информатизации по конференции телекоммуникационные и вычислительные системы. МТУСИ, Москва, 2008. – 140-141с.

3.Левинсон С.Е. Структурные методы автоматического распознавания речи // ТИИЭР, том 73, № 11, ноябрь, 1985. – с.100 - 128

4. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. - Киев: Наука. думка, 1987. – 264с

5. Ибрагимов Б.Г. Подход к реализации многофункциональных абонентских терминалов с использованием цифровой обработки сигналов // Автоматизация и современные технологий. №3, 2003, с.19-22.

ANALYSIS AND CHOICE OF SAMPLING FREQUENCY FOR PRELIMINARY PROCESSING OF SPEECH SIGNALS

Fatullaev A., Ibrahimov B.

Institute of Cybernetic of ANSA, Baku
Azerbaijan Technical University, Baku

In a present stage of development of systems of transfer and the processing of continuous signals at recognition of the speech information on the basis of perspective linguistic DSP (Digital Signal Processing) technologies get the large urgency in systems by a digital method of processing of speech signals. Thus the large interest causes of continuous speech signals on some entrance data, such as amplitude-frequency and amplitude-phase characteristics recognition of speech signals by means of transformation of an analog signal to a digital signal [1,2].

In systems of processing and recognition of speech the devices of input and transfer of continuous signals are used which realize preliminary processing of the entrance speech information with the purpose of reception more compact description of an entrance speech signal.

These systems get the important value in connection with introduction in practice low high-speed of digital channels communication, where link speed $V_k \leq 64$ Kbit/s.

In the given work the question of a choice of effective value of frequency sampling

for a speech signal in systems of recognition of speech is considered at processing continuous speech signals with application of a method sampling of processing of narrow-band signals, i.e. method undersampling of signals.

The mathematical formulation of the offered approach for continuous speech signals can be submitted by the following criterion function [3,4]: $E_д(\Delta F, t_{pc}, L_{дк}) = \{U(t_{pc}), U_d[kT_д], U_k[k\Delta t_{kv}], U[N(km)]\}$, k=0,1,2,… **(1)**
where $U(t_{pc})$ – function of amplitude of an entrance continuous speech signal t_c; $U_d[kT_д]$ – function sampling of an entrance signal on time as discrete readout; $U_k[k\Delta t_{kv}]$ – function of quantization entrance sampling of a signal on a level as discrete readout; U[N(km)] – function of binary calculation which is taking into account operation of coding sample of values of a transmitted signal as a sequence of binary code combinations with length $L_{дк}$; m – basis of a used code.

The expression (1) is analytical algorithm of transformation of a continuous signal, with which help the minimally effective value of possible frequency sampling $F_д$ the converter of a speech signal $U(t_{pc})$ on frequency to criteria can be determined.

For preliminary processing of a speech signal, on the basis of the carried out system-technical analysis, the decision realized on the basis of the structurally functional circuit of the digital processor of processing of signals with use of DSP-technology and is offered consisting from following blocks: the analyzer, quantizer and coder with the filter VAD (Voice Activate Detection).

In the considered block diagram the important module is the digital processor of processing of signals. His components are discreditizer, quantizer and coder with the filter, and also VAD.

Thus, as well as in a case sampling quadrate, the frequency sampling is limited from below to double width of a spectrum of a signal. In view of it from $(k+1)F_д - F_{max} > F_{max}$,

it is possible to define the greatest possible value k [1]: $k < \frac{2F_{min}}{F_Д} < \frac{F_{min}}{F_{max} - F_{min}}$, **(2)**

For all whole, not exceeding this value, from a double inequality $2F_{max} - F_д < k \cdot F_д < 2F_{min}$ **(3)**

is possible is to determined with a range of possible values of sampling frequency: $\frac{2F_{max}}{k+1} < F_Д < \frac{2F_{min}}{k}$ **(4)**

The choice of a step of quantization after a level of continuous signals approximations, providing process, of continuous signals by digital signals, that is, signals with discrete values of readout also is defined by the following inequality [5]: $\Delta t_{kv} \leq T_д$, $T_д = 2\pi / \omega_B$, $\omega_B \in [-\omega_B, +\omega_B]$ **(5),** where ω_B – width of the top frequency spectrum quainter of a continuous speech signal.

The results of research and analysis have shown, that received can be used for preliminary processing and input of the speech messages in system of speech recognition.

References

1. Sergienko A.B. Digital signal processing. Textbook for institute of higher education. 2- pupl. – SPb.: Piter, 2007. – 751c.

2. Fatullaev A.B., Ibrahimov B.G. The choice sampling frequency in systems speech recognition // Proceedings of International Academy Information's on the conference telecommunication and computer system. MTUCI, Moskow, 2008. – 140-142pp.
3. Levinson S.E. Structural methods in automatic speech recognition // Proceedings of the IEEE Trans. Informat. Theory. 1984, vol.73, N.11, pp.1625-1650.
4. Vinsuk T.K. Analyses, recognition and interpretations speech signals. – Kiev : Nauka, Dumka, 1987. – 264pp.
5. Ibrahimov B.G. The approach to realisation of multifunctional user's terminals with use of digital processing signals //Automation and modern technologies. No.3, 2003, pp.19-22.

УСТРОЙСТВО УНИВЕРСАЛЬНОЙ ПЕРЕПАКОВКИ ПОТОКОВ ДАННЫХ

Аминев Д.А., Батов А.А.

ЗАО "Московский научно-исследовательский телевизионный институт"

Введение

В системах сбора информации, системах регистрации и воспроизведения цифровых сигналов, системах спутникового приема и цифрового теле- радиовещания и т.п. возникает необходимость перепаковки потоков данных по нескольким каналам.

Известно множество решений по перепаковке потоков данных, но при обработке информации на компьютерах часто возникает задача ввода-вывода данных по нескольким каналам, причем их число может оперативно меняться, например, от одного до восьми.

Результаты исследований

Известны и широко используются такие устройства перепаковки потоков данных, как мультиплексоры и демультиплексоры, реализуемые как на специализированных интегральных схемах, так на ПЛИС.

В качестве аналогов можно привести микросхемы SN74LV164 и SN74LV165 фирмы Texas Instruments, которые позволяют выполнить перепаковку сигналов из одного потока в 8 и обратно. Фактически такая фиксированная схема перепаковки и является их недостатком.

Также имеются выделенные схемные компоненты ISERDES и OSERDES, входящие в состав ПЛИС семейства Virtex4 фирмы Xilinx и содержащие сдвиговые регистры. При этом компонент ISERDES предусматривает перепаковку данных из 1 потока в 2–8 потоков, а компонент OSERDES – перепаковку из 2–8 потоков в 1 поток. Эти схемные компоненты имеют следующие недостатки:

1) Такие компоненты принципиально требуют жесткой синхронизации входных и выходных потоков. То есть, для их работы каждый раз требуется две тактовые частоты, которые находятся в целочисленном соотношении между собой: для ISERDES выходная тактовая частота должна быть в 2–8 раз меньше входной (в зависимости от числа выходных каналов), а для OSERDES выходная тактовая частота должна быть соответственно в 2–8 раз больше входной (в зависимости от числа входных каналов).

2) Оба схемных компонента предназначены лишь для распараллеливания поступающего извне одного входного потока на несколько каналов внутри ПЛИС и для сбора нескольких потоков в один внутри ПЛИС для вывода во внешнюю цепь. В частности, они позволяют распараллеливать входной поток на четыре канала и собирать выходной поток из четырех каналов, что обеспечивает снижение требований по быстродействию для внутренней логики ПЛИС. Однако эти компоненты не обеспечивают решение задачи ввода и вывода нескольких параллельных потоков.

3) Физические ограничения по частотам ввода вывода в ПЛИС (поэтому такие компоненты целесообразно использовать для перепаковки из 4 потоков в 1 и обратно).

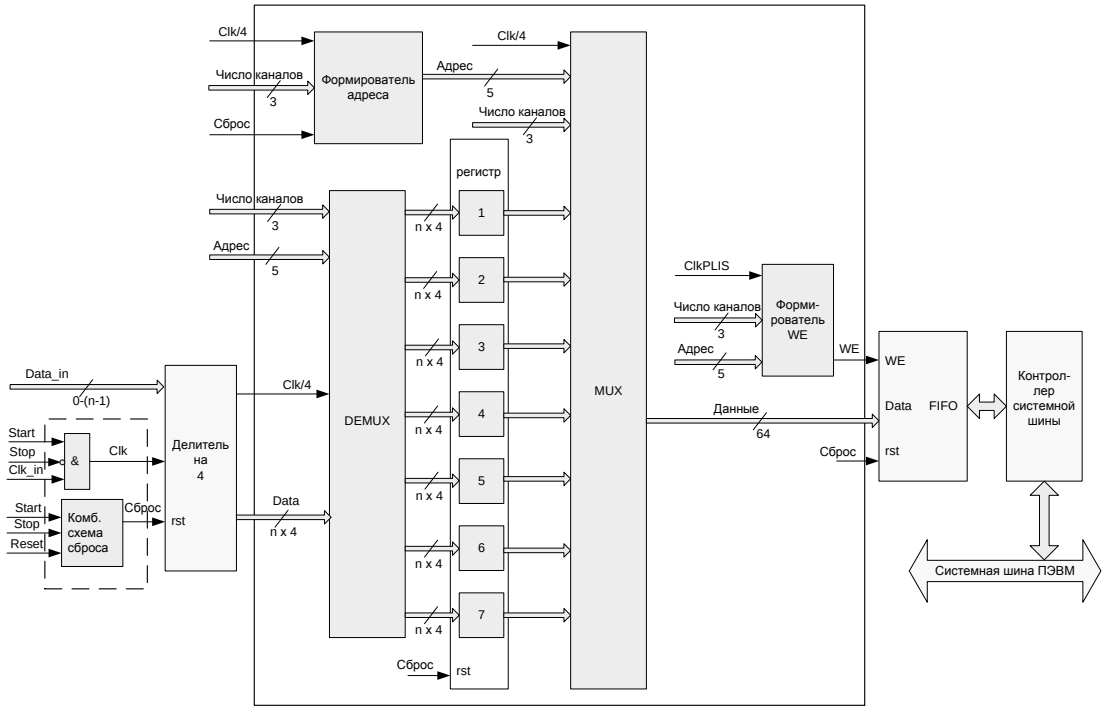

Рис. 1 – Структурная схема устройства – тракт ввода.

Описание устройства универсальной перепаковки потоков данных

Таким образом, учитывая недостатки существующих решений, предлагается устройство универсальной перепаковки потоков данных, состоящее из двух основных функциональных блоков – блок реализации тракта ввода и блок реализации тракта вывода потоков данных. Блок реализации тракта ввода осуществляет перепаковку данных из 1–8 входных потоков в 64-разрядный поток, используя входной тактовый сигнал синхронизации. Блок реализации тракта вывода осуществляет перепаковку данных из 64-разрядный потока в 1–8 выходных потоков и формирует выходной тактовый сигнал синхронизации.

На рис. 1 представлена структурная схема устройства – тракт ввода.

Описание входных сигналов

Data_in[n-1:0] – входные данные разрядностью n (от 1 до 8);

Start – сигнал старта (процесса приема потока данных);

Stop – сигнал остановки (процесса приема потока данных);

Clk_in – входной тактовый сигнал;

Reset – сигнал сброса, переводит все элементы схемы в начальное состояние;

Число каналов [2:0] – код числа каналов ввода (от 1 до 8).

Описание внутренних сигналов

Clk/4 – тактовый сигнал с частотой в четыре раза меньше входного

тактового сигнала Clk_in;

ClkPlis – тактовый сигнал контроллера системной шины;

Data[n*4] – данные разрядностью n*4, полученные путем распараллеливания входных данных Data_in[n-1:0];

Адрес[4:0] – адресная шина;

Описание выходных сигналов

Сброс – преобразованный сигнал сброса Reset;

WE – сигнал разрешения записи для FIFO;

Data[63:0] – 64-разрядные данные записываемые в FIFO.

Основная задача устройства на тракте ввода – преобразовать входной n-разрядный поток данных (n = 1, 2 … 8) в 64-разрядный и записать его в FIFO.

Делитель на 4 используется для деления частоты приходящих данных на 4 и распараллеливания каждого входного канала на 4 потока.

Демультиплексор DEMUX производит запись данных по n-разрядной шине в разрядный регистр, а мультиплексор MUX производит считывание данных из регистра.

Формирователь адреса задает адрес для мультиплексоров, в соответствии с которым будет произведена запись в требуемые разряды регистра. Формирователь WE предназначен для формирования сигнала разрешения записи данных в FIFO.

Контроллер системной шины передает данные из FIFO в компьютер.

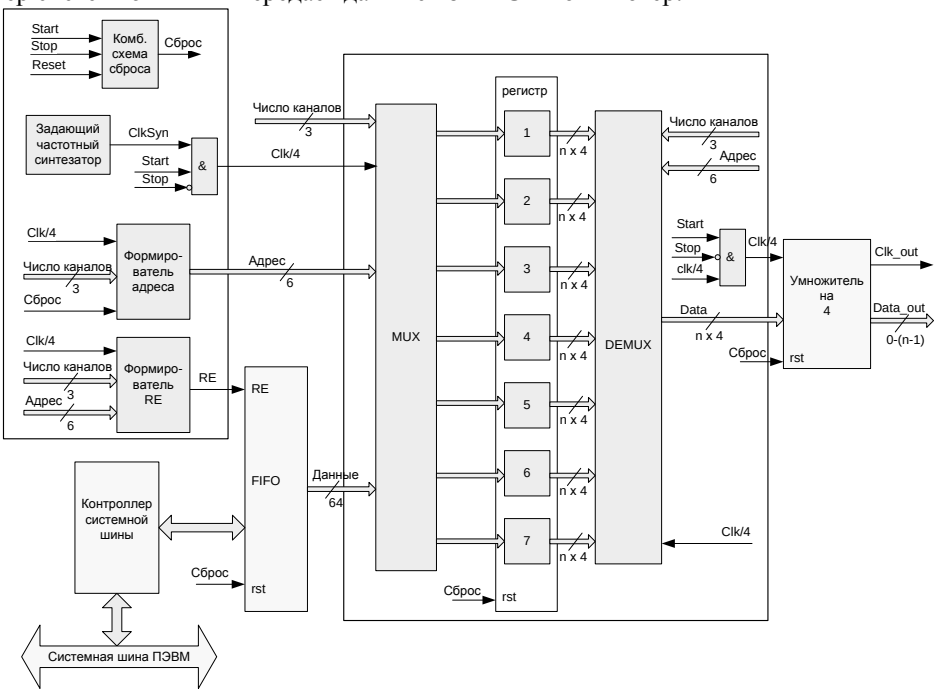

Рис. 2 – Структурная схема устройства – тракт вывода.

На рис. 2 представлена структурная схема устройства – тракт вывода.

Описание входных сигналов:

Data[63:0] – 64-разрядные данные читаемые из FIFO;
Start – сигнал старта (процесса приема потока данных);
Stop – сигнал остановки (процесса приема потока данных);
Reset – сигнал сброса, переводит все элементы схемы в начальное состояние;
Число каналов [2:0] – код числа каналов вывода (от 1 до 8).
Описание внутренних сигналов
ClkSyn – тактовый сигнал задающего частотного синтезатора;
Clk/4 – тактовый сигнал с частотой в четыре раза меньше выходного
тактового сигнала Clk_out;
Data[n*4] – данные разрядностью n*4 читаемые из регистра;
Адрес[5:0] – адресная шина;
Описание выходных сигналов
Сброс – преобразованный сигнал сброса Reset;
RE – сигнал разрешения чтения из FIFO;
Clk_out – выходной тактовый сигнал;
Data[n-1:0] – выходные данные разрядностью n (от 1 до 8);

Основная задача устройства на тракте вывода – преобразовать входной 64-разрядный поток из FIFO данных в n-разрядный (n = 1, 2 … 8).

Контроллер системной осуществляет передачу данных из компьютера в FIFO.

Формирователь RE предназначен для формирования сигнала разрешения чтения 64 бит данных из FIFO .

Формирователь адреса задает адрес для мультиплексоров, в соответствии с которым будет произведена запись в требуемые разряды регистра.

Мультиплексор MUX производит запись данных по 64-разрядной шине в регистр, а мультиплексор DEMUX производит считывание данных из регистра.

Умножитель на 4 используется для умножения частоты выходных данных на 4 и уплотнения n*4-разрядного потока в n-разрядный поток.

Временные диаграммы работы устройства

Представленные на рис. 3 временные диаграммы отображают работу устройства в режиме соединения выходов блока реализации тракта вывода с входами блока реализации тракта ввода.

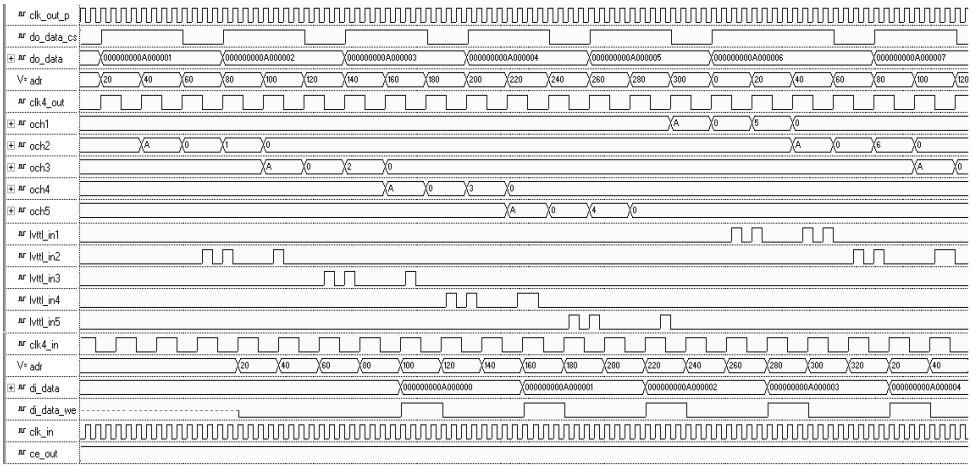

Рис. 3 – Временные диаграммы работы устройства при пяти каналах ввода/вывода.

По временным диаграммам видно, что входная возрастающая последовательность 64-разрядных данных, проходя через тракт вывода устройства, преобразуется в 5-разрядный поток, который поступает на тракт ввода устройства и снова преобразуется в последовательность 64-разрядных данных. Входная возрастающая последовательность на входе соответствует последовательности на выходе.

Заключение

Устройство универсальной перепаковки потоков данных может быть реализовано на отдельных цифровых микросхемах, программируемых логических интегральных схемах и на специализированных логических интегральных схемах.

Литература

1. Virtex-4 Libraries Guide for HDL Designs. – Xilinx, 2005.
2. Калабеков Б.А.. Цифровые устройства и микропроцессорные системы. М.: Горячая линия – Телеком, 2003
3. Low-Voltage Logic Data Book. – Texas Instruments, 1996.

UNIVERSAL DEVICE FOR REPACK DATA STREAMS

Aminev D., Batov A.

“Moscow Scientific - Research Television Institute”

In the data processing on PC often appear a problem of the input-output data on the few homogeneous channels and the number of it can be changed from one to eight, for example.

In the usually case that logical task of the data repack can be resolved by the data join in the single continuous stream (with the appropriate frequency increment) and then perform division on a number of the parallel streams. That streams is defined the system bus construction. But that similar structure can be easily realized for the low-speed input streams, but for the extremely high-speed the integrated streams is difficult realized. The output data task is reversed but the all speed limits is such as in input data task.

The idea for the entity is the change of the classical synchronous data conversion circuit on the asynchronous circuit. It permit theoretically implement the input-output data on the unlimited speed. The speed is limited only by technology limits. The operative change of the number of the channels is provided by choice of the intermediate buffers capacity.

The writing in the buffer performs by the bit groups on the data input. The number of the bit groups is conforming to the number of input channels. The reading from the buffer perform parallel on the all channels. That channels is defined by system bus architecture. Analogue approach may be used on the data output.

That universal device for repack data streams may be realized in the small integrated circuits, programmable logic devices (for example Xilinx XC4VFX2011 series Virtex 4) and application-specific integrated circuits.

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

Веселов И.А., Новосёлов С.А., Новиков А.Е., Топников А.И.

Ярославский государственный университет имени П.Г. Демидова

Введение

Распознавание речи, одного из важнейших способов человеческой коммуникации, является значительной частью задачи усовершенствования интерфейсов между человеком и компьютером. Под

распознаванием речи может пониматься преобразование речи в текст, распознавание и выполнение определенных команд, обработка и извлечение каких-либо ключевых параметров. В работе затрагивается проблема распознавания голосовых команд. Несмотря на очевидный прогресс в данной области исследований, распознавание речи продолжает оставаться сложной проблемой. Уже существуют эффективные алгоритмы распознавания голосовых команд, однако, до сих пор актуален вопрос о методе и приёмах, которые могут использоваться для решения поставленной задачи. Процедура распознавания голосовой команды состоит из двух основных этапов: этап предварительной обработки, фильтрации и выделения ключевых информативных параметров речи и этап непосредственного сравнения входящей реализации команды с множеством заранее созданных реализаций эталонов. Предлагается возможность применения аппарата вейвлет-анализа для реализации первого этапа и аппарата скрытых Марковских моделей для реализации второго.

Скрытые Марковские модели

Рассмотрим систему, которая в произвольный момент времени может находиться в одном из N различных состояний S_1, S_2, …, S_N. В дискретные моменты времени система претерпевает изменение состояния (возможно, переходя при этом опять в то же состояние) в соответствии с некоторым вероятностным правилом, связанным только с текущим состоянием. В каждом таком состоянии система в соответствии уже с другим вероятностным правилом выдает символ наблюдения, один из M возможных V_1, V_2, …, V_M. Для полного вероятностного описания такой системы необходимо задать три матрицы вероятностей:

1. Начальное распределение вероятностей состояний $\pi = \{\pi_i\}$, где π_i – вероятность того, что в начальный момент времени находится в состоянии S_i.

2. Распределение вероятностей переходов между состояниями (или матрица переходных вероятностей) $A = \{a_{ij}\}$, где a_{ij} – вероятность того, что из состояния S_i система перейдёт в состояние S_j.

3. Распределение вероятностей появления символов наблюдения $B = \{b_j(k)\}$, где $b_j(k)$ – вероятность того, что в состоянии S_j будет выдан символ наблюдения V_k.

Такую систему называют скрытой Марковской моделью и обозначают как $\lambda = \lambda(A,B,\pi)$. Её результатом будут две последовательности: состояний **S** (которая скрыта и в данном случае интересовать не будет) и наблюдений **O** (которая состоит из символов наблюдений V).

В теории скрытых Марковских моделей [1,2] существуют три основные задачи. В работе использованы решения двух из них.

Первая задача. Пусть заданы последовательность наблюдений $O = \{O_t\}$ и модель $\lambda = \lambda(A,B,\pi)$. Необходимо вычислить вероятность появления этой последовательности наблюдений для данной модели, т.е. найти $P(O|\lambda)$. Это обычная задача оценивания.

Вторая задача. Пусть заданы последовательность наблюдений $O = \{O_t\}$ и модель $\lambda = \lambda(A,B,\pi)$. Каким образом нужно подстроить параметры модели (изменить A, B и π), чтобы максимизировать $P(O|\lambda)$? Задача является оптимизационной, с её помощью «обучают» модель. Итеративно производя подстройку модели, можно добиться желаемого качества её соответствия последовательности наблюдений.

Обе задачи имеют аналитические решения, которые можно найти, например, в [1].

Решение второй задачи можно распространить на случай нескольких последовательностей наблюдений $\mathbf{O}_1$, $\mathbf{O}_2$, …, $\mathbf{O}_n$. Тогда будет максимизироваться произведение вероятностей появления отдельных последовательностей $P(O_1|\lambda) * P(O_2|\lambda) * ... * P(O_n|\lambda)$. При распознавании речи это позволяет строить одну модель для нескольких дикторов.

В работе скрытые Марковские модели использовались в качестве классификатора.

Входными параметрами для моделей являлись энергии полос вейвлет-разложения.

Речевой сигнал является примером нестационарного процесса, в котором информативным является сам факт изменения его частотно-временных характеристик. Для выполнения анализа таких процессов требуются базисные функции, обладающие способностью выявлять в анализируемом сигнале как частотные, так и его временные характеристики. Другими словами, сами функции должны обладать свойствами частотно-временной локализации.

Идея дискретного вейвлет-анализа [3] состоит в представлении сигнала последовательностью образов с разной степенью детализации (многомасштабный анализ), что позволяет выявлять локальные особенности сигнала и классифицировать их по интенсивности. Как показано на рис. 1, дискретное симметричное вейвлет-преобразование осуществляется с использованием цифровых низкочастотного и высокочастотного вейвлет-фильтров G и H и блоков децимации. В процессе разложения участвуют вейвлеты Добеши (Daubechies) – db1, db2, db3, …, где последняя цифра обозначает количество нулевых моментов.

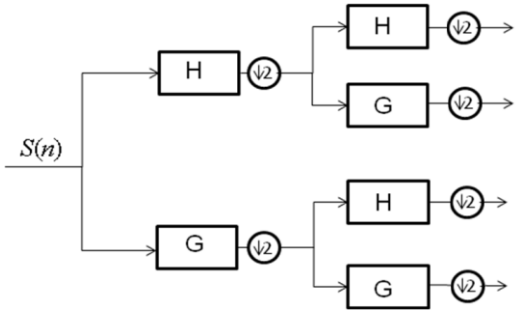

Рис. 1. Дискретное симметричное вейвлет-разложение

Алгоритм

Имеется обучающая база тестовых речевых команд.

1. Для каждого сигнала из базы применяется симметричное вейвлет-разложение и считается энергия вейвлет-коэффициентов в каждой полосе. В результате получается последовательность из 2^N параметров (где N – уровень разложения). Это последовательность квантуется.
2. По последовательностям, полученным для нескольких образцов одной команды, строится скрытая Марковская модель, соответствующая этому сигналу. При обучении осуществляется определенное количество итераций, которое влияет на правильность распознавания. В результате получаем множество моделей (по одному для каждой речевой команды из базы).
3. Неизвестный сигнал, который необходимо распознать, подвергается такой же обработке, как и команды из базы в пункте 1. В результате получается «неизвестная» последовательность параметров.
4. Эта последовательность подается на все созданные модели. Вычисляется вероятность её появления для каждой модели (путём решения первой задачи). Делается вывод, что неизвестный сигнал совпадает с той командой из базы, для модели которой получена максимальная вероятность. На практике возникали ситуации, когда максимум однозначно определить не удаётся (например, все вероятности нулевые), в этом случае делается вывод, что неизвестный сигнал не соответствует ни одной команде.

Исследования

База состояла из цифр, произнесённых на русском языке одним диктором, по 30 образцов для каждой. Все сигналы выровнены по энергии. Построено десять Марковских моделей, каждая для 30 сигналов. Распознавались только цифры, произнесённые тем же диктором. Алгоритм был реализован в среде программирования MatLab.

Представляло интерес выявить зависимость процента правильно распознанных цифр от:

1. уровня вейвлет-разложения;
2. количества итераций при обучении модели;
3. гладкости вейвлета, используемого при разложении.

Результаты приведены в табл. 1 и табл. 2.

Таблица 1. Вероятность распознавания для уровней разложения 3 и 4, вейвлетов Добеши 2,3 и 6 и количества итераций от 10 до 80

	уровень разложения-3			уровень разложения-4		
количество итераций	db2	db3	db6	db2	db3	db6
10	0,85	0,83	0,65	0,80	0,80	0,75
20	0,85	0,83	0,65	0,80	0,78	0,73
30	0,85	0,82	0,65	0,80	0,78	0,73
40	0,85	0,82	0,65	0,80	0,78	0,72
50	0,85	0,82	0,65	0,80	0,78	0,72
60	0,85	0,82	0,65	0,80	0,78	0,72
70	0,85	0,82	0,65	0,75	0,78	0,72
80	0,85	0,82	0,65	0,75	0,78	0,72

Таблица 2. Вероятность распознавания для уровней разложения 5 и 6, вейвлетов Добеши 2, 3 и 6 и количества итераций от 10 до 80

	уровень разложения-5			уровень разложения-6		
количество итераций	db2	db3	db6	db2	db3	db6
10	0,70	0,82	0,78	0,38	0,45	0,70
20	0,70	0,85	0,87	0,52	0,50	0,70

30	0,72	0,85	0,90	0,55	0,53	0,70
40	0,72	0,85	0,90	0,55	0,53	0,70
50	0,70	0,85	0,90	0,47	0,55	0,78
60	0,73	0,85	0,90	0,47	0,60	0,77
70	0,73	0,85	0,90	0,47	0,65	0,77
80	0,73	0,85	0,90	0,47	0,65	0,77

Выводы

Предложен алгоритм распознавания речевых команд при использовании информативных параметров, основанных на вейвлет-разложении сигнала, и скрытых Марковских моделей в качестве классификатора. Установлено, что оптимальным с точки зрения вероятности правильного распознавания является пятый уровень разложения. Более низкие уровни (2, 3, 4) дают меньшую вероятность из-за того, что Марковские модели путают сигналы друг с другом, так как такой крупный масштаб не позволяет выявить отличительные особенности каждой команды. Более высокие уровни (6, 7) так же уменьшают вероятность распознавания, потому что возрастает количество ситуаций, когда система делает неправильный вывод об отсутствии входной команды в базе. Это происходит из-за того, что модель подстраивается под достаточно мелкий масштаб, в то время как реальные образцы одной команды могут отличаться друг от друга сильнее. Решение этой проблемы видится в увеличении обучающей базы команд. Оптимальным количеством итераций является 50-60. При меньшем значении система ещё не достаточно хорошо подстроилась под команду. Дальнейшее увеличение количества итераций лишь увеличивает вычислительную сложность программы, при этом мало влияя на вероятность. Наилучшим с точки зрения гладкости является вейвлет Добеши 6.

Полученные результаты для вероятности распознавания согласуются с другими алгоритмами, использующими вейвлет-параметры сигнала. Однако пока разработанный алгоритм несколько уступает кепстральным методам.

Исследования данного алгоритма предполагается продолжить. В частности, провести эксперименты для дикторонезависимого случая. Так же предлагается использовать в качестве информативных параметров команд энергии полос вейвлет-разложения сразу нескольких уровней. В перспективе планируется создание самообучающейся системы, то есть в процессе работы система сама будет изменять параметры скрытых Марковских моделей, подстраивая их под новые входные данные.

Литература

1. Рабинер Л.Р. Скрытые Марковские модели и их применение в избранных приложениях при распознавании речи: Обзор. // ТИИЭР, 1989. Т. 77, № 2.
2. Benesty, Sondhi, Huang (Eds.) Springer Handbook of Speech Processing. // Springer 2008.
3. Daubechies I. Ten Lectures on Wavelets. SIAM, Philadelphia, PA, 1992.

THE APPLICATION OF WAVELET TRANSFORM AND HIDDEN MARKOVS MODELS IN THE SPEECH COMMANDS RECOGNITION PROBLEM

Veselov I., Novoselov S., Novikov A., Topnikov A.
Yaroslavl State University

The speech recognition one of the major ways of the human communications. It is a significant part of improvement of interfaces between the person and a computer. The speech to the text transformation, commands recognition can be understood as the speech recognition. In the given work the problem of the voice commands recognition is considered. Despite of obvious progress in the given area of researches, the speech recognition remains a complex problem. Already there are effective algorithms of voice commands recognition ,however, the question on a method and receptions which till now is actual can be used for the solving of a task in view. Procedure of recognition of a voice command consists of two basic stages: Predesign stage, filtrations and allocation of key informative parameters of speech and a stage of direct comparison of entering realization of a command with set of beforehand created realizations of standards. The opportunity of application of the device Wavelet-analysis for realization of the first stage and the device hidden Markovs models for realization of the second is offered [1].

The speech signal is an example of non-stationary process in which the fact of change of its time-and-frequency characteristics is informative. To the analysis of speech signals pertinently to apply such mathematical method as wavelet - transformation.

Algorithm

There is training base of test speech commands.

1. Symmetric wavelet-decomposition is applied to each signal from base and energy wavelet-factors in each strip is considered. In result the sequence from 2^N parameters (where N - a level of decomposition) turns out. It is a sequence квантуется.
2. On the sequences received for several samples of one command, is under construction latent Markovs model appropriate to this signal. At training the certain quantity of iterations which influences correctness of recognition is carried out. In result we receive set of models (on one for each speech command from base.

3. The unknown signal which is necessary for distinguishing, is exposed to the same processing, as well as commands from base in item 1. In result the "unknown" sequence of parameters turns out.
4. This sequence is moved to all created models. The probability of its occurrence for each model (is calculated by the solving of the first task). It is judged, that the unknown signal coincides with that command from base for which model the maximal probability is received. In practice there were situations, when a maximum is unequivocal to define not possible (for example, all probabilities zero), in this case is judged, that the unknown signal does not meet to any command.

References

1. Benesty, Sondhi, Huang (Eds.) Springer Handbook of Speech Processing. // Springer 2008.

СИСТЕМА ПОИСКА КЛЮЧЕВЫХ СЛОВ В НЕПРЕРЫВНОМ РЕЧЕВОМ ПОТОКЕ

Гладышев К.К.

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича

Одной из актуальных задач в области речевых технологий, является поиск определенных слов в потоке разговорной речи. Набор таких слов, как правило, ограничен. Необходимо определить, встречаются ли данные слова в произнесенных фразах, и зафиксировать время начала и окончания их звучания.

Автором статьи разработана экспериментальная система по распознаванию ключевых слов или целых фраз в непрерывном речевом потоке (слитной речи). Система является иерархической, основана на бионической модели восприятия речи человеком [3] и состоит из нескольких взаимосвязанных модулей.

Обрабатываемые речевые сигналы подаются на вход системы в оцифрованном виде. Данная операция выполняется с помощью микрофона и звуковой карты ПК. Очевидно, что использовать представление звука во временной форме для задач распознавания речи неэффективно, т.к. оно не отражает характерных особенностей звукового сигнала. Необходимо наличие блока по выделению эффективных информативных признаков речевого сигнала. К настоящему времени известны различные варианты моделей и методов выделения акустических признаков речевых сигналов. В разработанной системе используется аппарат линейного предсказания [2]. Получаемые признаки – линейные спектральные корни (ЛСК), обладают рядом полезных свойств – они просто рассчитываются, дают компактное представление речевых сигналов, наименее чувствительны к действиям помех и смене диктора. Исходный сигнал разбивается на отрезки (окна или кадры) определенной длины. Кадры перекрываются между собой. На каждом кадре производится расчет набора ЛСК. В результате речевой сигнал представляется в виде массива точек в многомерном пространстве признаков ЛСК.

На первой стадии необходимо провести обучение системы. Диктором записывается набор эталонных речевых единиц (например, слов), поиск которых необходимо будет проводить. Для всех элементов производится расчет наборов ЛСК, данные сохраняются в базе. Система обучена и готова к распознаванию.

Для обеспечения работы системы в режиме реального времени используется накопительный буфер, который позволяет сохранять отрезки сигнала определенной длительности. За это время производится распознавание предыдущего речевого фрагмента. Таким образом, дальнейшая обработка производится на сигналах конечной длительности. Размер буфера равен средней длительности звучания эталонов из словаря. Для каждого фрагмента входного речевого сигнала производится расчет набора ЛСК.

На следующем этапе для анализируемого речевого фрагмента необходимо провести поиск ближайшего представителя по словарю. Выполняется последовательное сравнение с каждым из эталонов с помощью динамической свертки (или динамического программирования) [1]. Подсчитывается минимальное накопленное расстояние при переходе системы из состояния, соответствующего набору ЛСК одного сигнала, в состояние, соответствующее другому образцу речевого сигнала. При этом учитывается временная последовательность ЛСК. На выходе процедуры сравнения получается некоторое число (мера близости). Чем оно больше, тем более различаются эталон и входной сигнал. В качестве меры расстояния между многомерными векторами сигналов используется Евклидова метрика.

Одним из основных преимуществ динамической свертки является автоматическое масштабирование во временной области для различных по длительности образцов. В случае речевых сигналов, нет необходимости точной подгонки длительности сигналов, четкого вырезания пауз и т.д. Важно, что темп произнесения слов может быть разным, например, отдельные гласные могут тянуться человеком.

Распознанным эталоном на текущем кадре речевого сигнала будет являться тот, до которого подсчитано минимальное накопленное расстояние. Если мера близости превышает определенный порог, значит, на текущем кадре не встречается искомых ключевых слов. Величина порога определяется экспериментально и является настроечным параметром системы.

Временная последовательность анализируемых кадров соотносится с входящим РС. Благодаря этому по результатам сравнения с эталонами определяются границы искомых слов в анализируемой фразе.

На рисунке 1 представлена временная диаграмма фразы «черная тойта номер три два один в сторону Питера». На рисунках 2-3 показаны результаты поиска различных слов в данной фразе. По горизонтальной оси отложены номера кадров, на которые разбивается входящий речевой сигнал, по вертикальной оси значения меры близости до искомого эталона. Видно, что в обоих случаях для искомых слов наблюдаются минимумы. Это свидетельствует об успешности разработанной системы.

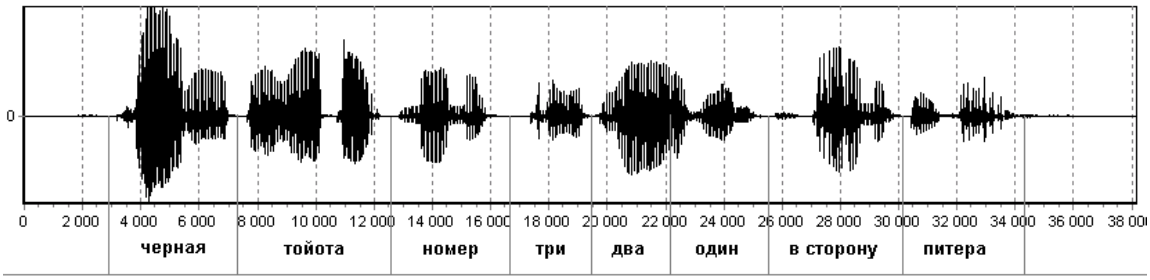

Рис 1. Временная диаграмма фразы.

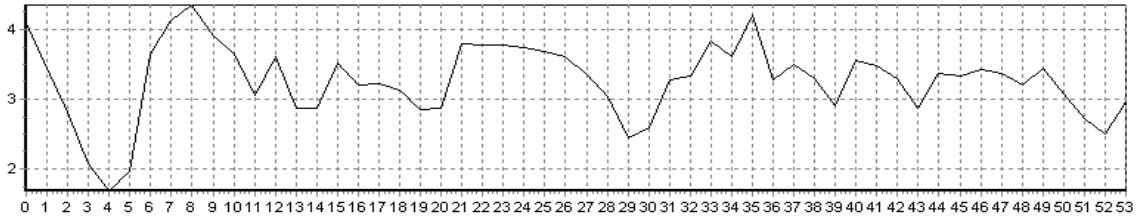

Рис 2. Поиск слова «черная» во фразе.

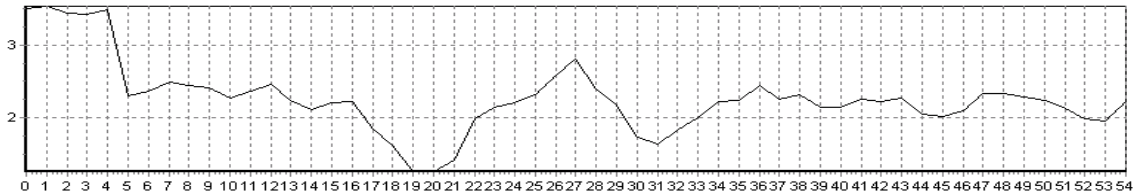

Рис 3. Поиск слова «номер» во фразе.