

УДК 004.942

И.Е. Щеголев

## Оценка времени выполнения и погрешности обработки речевого сигнала при использовании модели слуха человека

В статье рассматриваются результаты экспериментов по тестированию модели слуховой системы на различных программных и аппаратных средствах. Приводятся рекомендации по использованию различных типов данных и архитектур процессора.

Обработка и исследование речевого сигнала позволяет решить широкий круг задач, среди которых можно отметить: моделирование слухового восприятия речевого сигнала, исследование структуры и параметров речевого сигнала, исследование процессов сегментации сигнала по признаку наличия голосового источника [1–3]. В настоящее время существующее алгоритмическое и программное обеспечение не удовлетворяет ряду требований для решения практических задач. В первую очередь это требования по быстродействию и точности вычисления параметров. В работах [4–5] приведены основы модели слуховой системы человека и реализация данной модели в виде программного комплекса, на основе которых была выполнена оценка времени выполнения и погрешности обработки речевого сигнала.

Очевидно, что эффективная реализация алгоритмов строится на полнофункциональном тестировании всех критических секций по времени выполнения и допустимой погрешности. Для этого были проведены исследования быстродействия и точности алгоритмов обработки речевых сигналов, зависимости точности определения основных параметров сигнала от характеристик, влияющих на быстродействие.

Исследование точности и быстродействия алгоритмов проводилось на речевых и синтезированных сигналах с частотой дискретизации 8 кГц, записанных в 16-битной цифровой форме. В качестве речевого сигнала была взята цифровая запись фразы «Эти жирные сазаны ушли под палубу», произнесенной женским голосом. Данная фраза взята из набора тестовых фраз для оценки качества речи и узнаваемости голоса диктора [6]. Длительность звучания фразы составляет 2,41 секунды. Для экспериментов были использованы банки фильтров с количеством каналов фильтрации, равным 64, 100, 128, 256, 350, 512, 800 и 1024. Применялись следующие типы данных для коэффициентов: вещественный размером 4 байта, вещественный размером 8 байт, вещественный размером 10 байт. Были исследованы алгоритмы с учетом краевого эффекта и без.

Из всех рассмотренных алгоритмов быстродействие измерялось только для алгоритма свертки как самого емкого по времени участка. На рисунке 1 представлены графики зависимости времени выполнения алгоритма свертки от количества каналов фильтрации (на рисунке 1 и далее «кэ» – учет краевого эффекта).

Очевидно, что с ростом размера типа данных коэффициентов практически пропорционально растет время работы алгоритма. Данные замеры проводились на 32-разрядном процессоре AMD Athlon XP 1800+ с тактовой частотой 1,5 ГГц. На двух других 64-разрядных процессорах AMD Athlon 64 3200+ (тактовая частота 2,0 ГГц) и AMD Opteron 2216 (тактовая частота 2,4 ГГц) по сравнению с первым процессором наблюдается уменьшение скорости обработки данных типа «вещественный размером 8 байт» практически до уровня скорости обработки данных типа «вещественный размером 4 байта». На рисунке 2 представлены аналогичные графики для процессора AMD Opteron 2216.

По полученным данным можно сформулировать следующие рекомендации. При реализации исследуемых алгоритмов на процессорах, схожих по архитектуре с AMD Athlon 64 3200+ и AMD Opteron 2216 использование типа данных «вещественный 4 байта» по сравнению с типом «вещественный 8 байт» не дает преимуществ в быстродействии, но в два раза увеличивает объем хранимых и обрабатываемых данных.

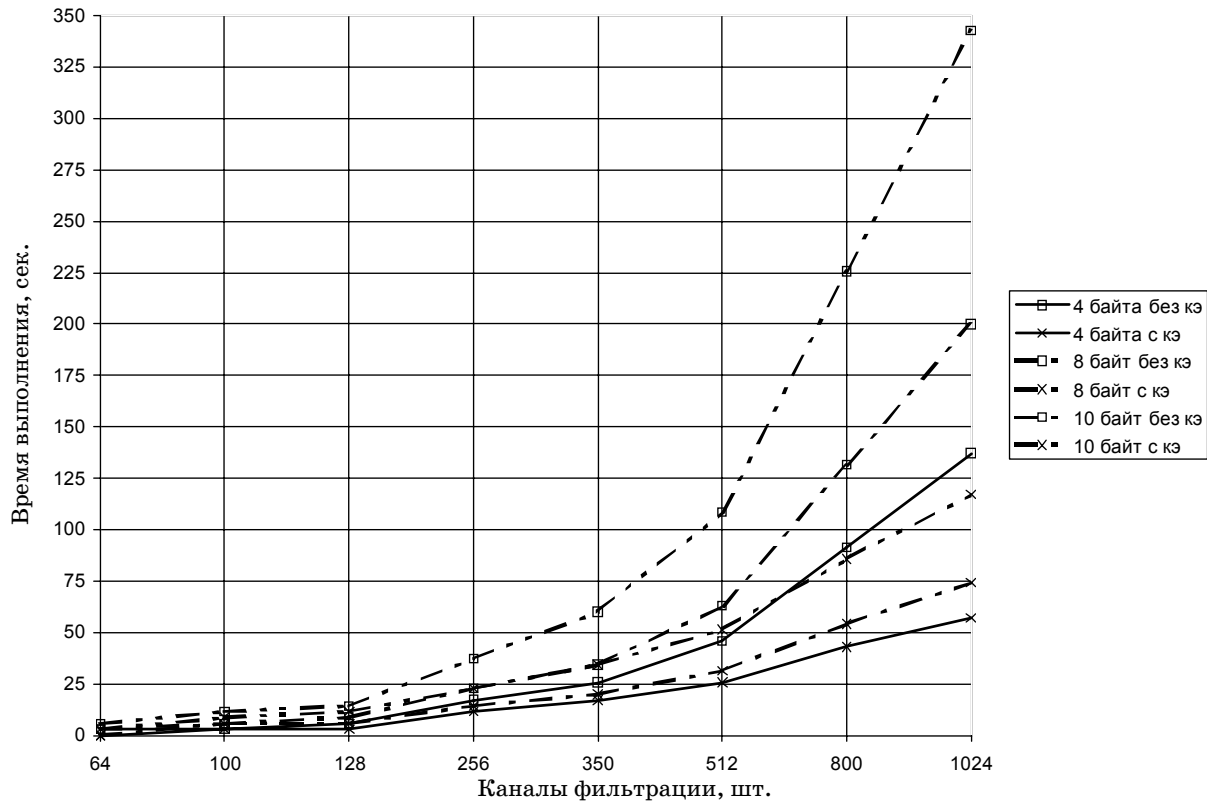


Рис. 1. Графики зависимости времени свертки от количества каналов фильтрации (AMD Athlon XP 1800+)

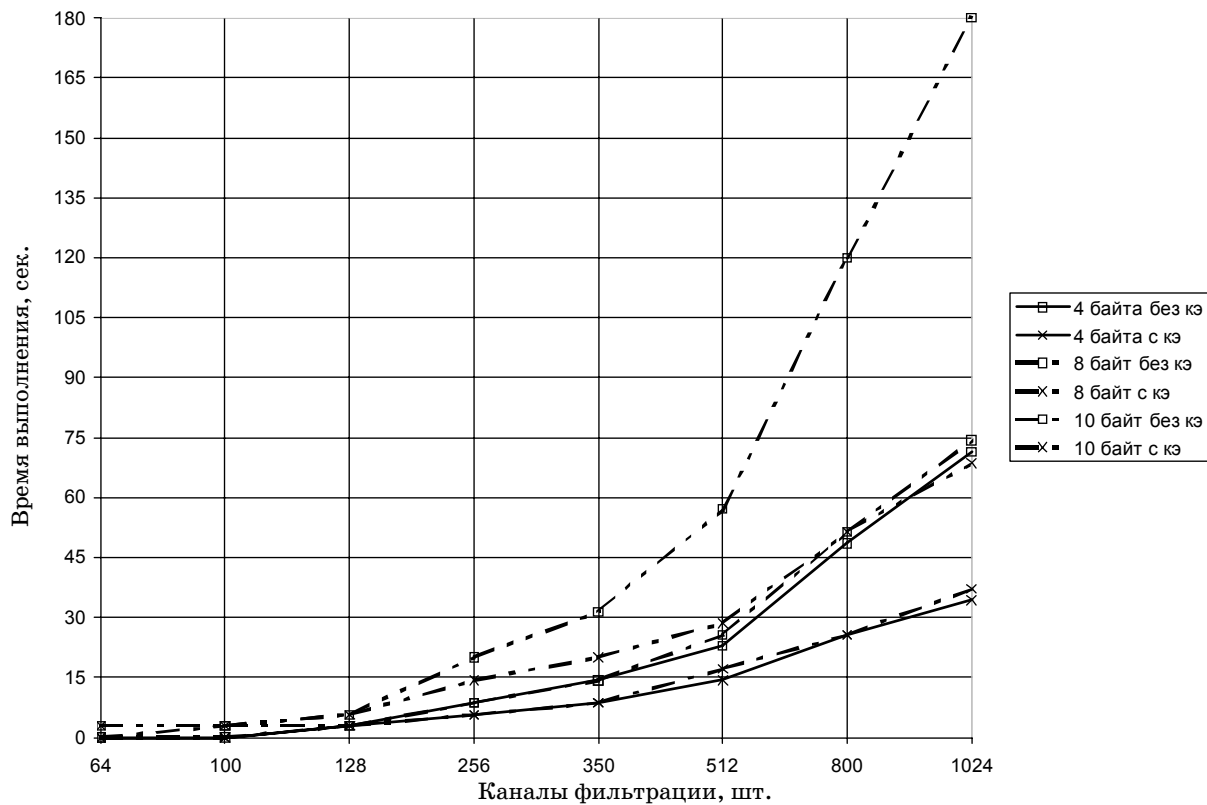


Рис. 2. График зависимости времени свертки от количества каналов фильтрации (AMD Opteron 2216)

Для измерения точности на речевом сигнале предварительно были определены локализованные участки и выбраны 82 отсчета на этих участках.

В качестве эталонных значений, относительно которых измеряется погрешность вычислений, были выбраны значения частоты и интенсивности основного тона на выбранных отсчетах при 1024 каналах фильтрации. В ходе эксперимента понижалось число каналов и менялся тип данных для коэффициентов.

На рисунке 3 представлены графики зависимости погрешности вычисления частоты основного тона от количества каналов фильтрации. С ростом количества каналов наблюдается уменьшение величины погрешности вычислений. Реализации с применением различных типов данных дают близкие значения погрешности (отличия в шестой значимой цифре числа). Реализации с учетом краевого эффекта и без его учета отличаются друг от друга на сотые доли процента.

На рисунке 4 представлены графики зависимости погрешности вычисления интенсивности основного тона от количества каналов фильтрации. С ростом числа каналов наблюдается уменьшение значения погрешности вычислений. Реализации с применением различных типов данных дают очень близкие значения погрешности (отличия в пятом знаке после запятой). Реализации с учетом краевого эффекта и без его учета отличаются друг от друга не более чем на 5 процентов.

В качестве синтезированных сигналов при измерении точности вычислений частоты были взяты сигналы синусоидальной формы с десятью гармониками и частотой от 100 до 600 Герц с шагом в 50 Герц. Длительность звучания была выбрана равной 100 миллисекундам для каждого синтезированного сигнала, что позволило существенно сократить время проведения экспериментов.

В качестве эталонного значения, относительно которого измеряется погрешность вычислений, выбрана частота первой гармоники исходного сигнала, известная а priori. В ходе эксперимента изменялись число каналов и тип данных для коэффициентов.

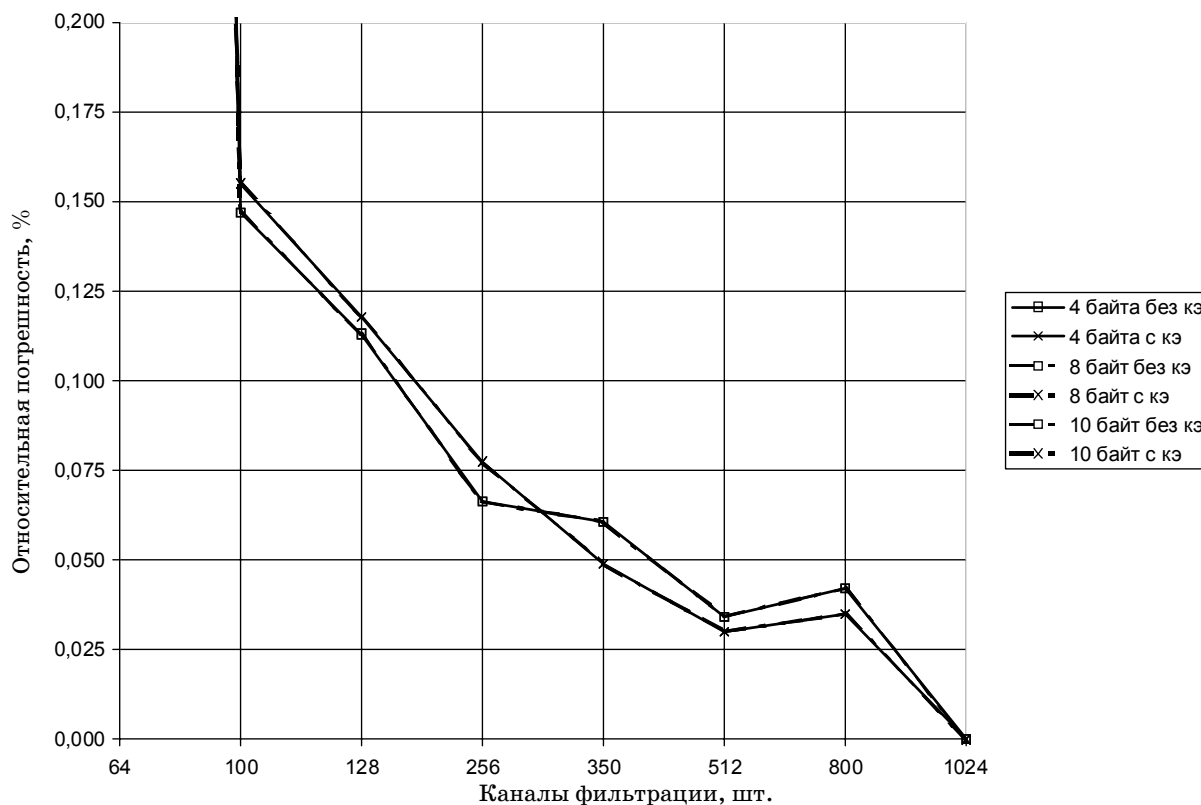


Рис. 3. Зависимость относительной погрешности вычисления частоты основного тона от количества каналов фильтрации

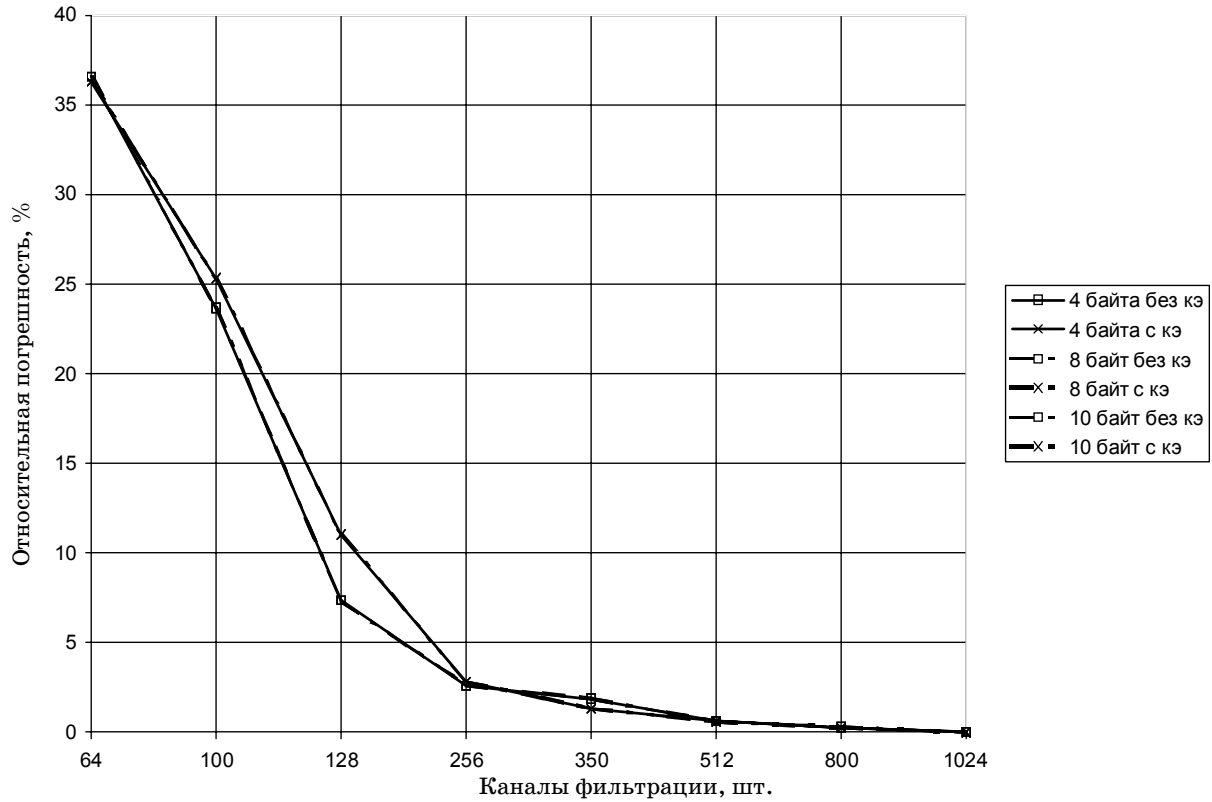


Рис. 4. Зависимость относительной погрешности вычисления интенсивности основного тона от количества каналов фильтрации

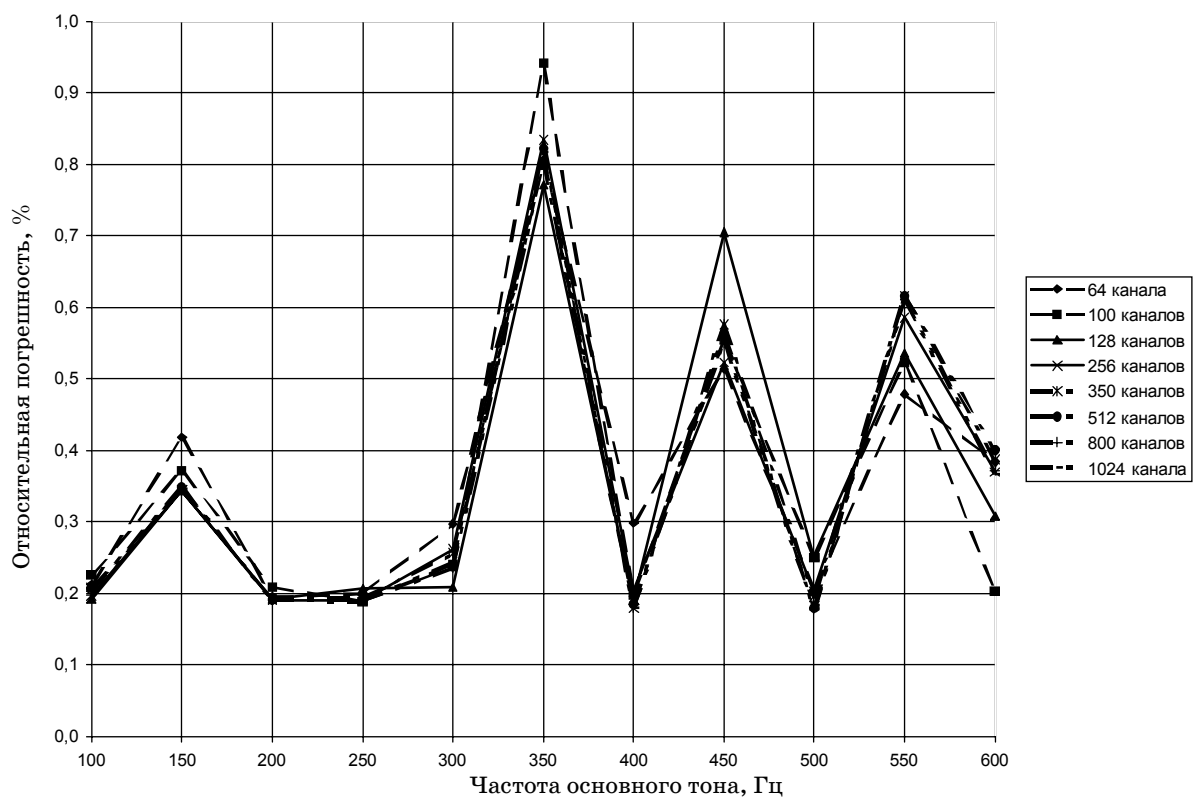


Рис. 5. Зависимость относительной погрешности вычисления частоты основного тона от частоты основного тона сигнала

По результатам исследования зависимости погрешности вычисления частоты основного тона от количества каналов фильтрации можно сделать вывод, что максимальная погрешность вычисления частоты составляет 5 Герц или 1 процент. С изменением числа каналов погрешность изменяется на десятые доли процента. Характер зависимости частоты от количества каналов не прослеживается.

На рисунке 5 представлены графики зависимости погрешности вычисления частоты основного тона от заданной частоты основного тона синтезированного сигнала (реализации без учета краевого эффекта для типа данных вещественный размером 4 байта).

Из графиков следует, что алгоритм дает погрешность не более чем 1 процент. Причем экстремумы зависимости, отчетливо проявляющиеся на частотах 350, 450 и 550 Герц, можно объяснить погрешностью генератора простейших сигналов. Другими словами, на вход алгоритма свертки был подан сигнал с частотой, отличной от номинально заданной при его синтезе.

Сформулируем рекомендации по подбору параметров алгоритмов обработки речевого сигнала. Наборы фильтров с числом каналов, равным 128, хорошо подходят для исследований, не требующих большой точности. Данный вариант обеспечивает погрешность вычисления частоты речевого сигнала не более чем 0,12 процента, погрешность вычисления интенсивности при этом не превышает 11,02 процента. Для исследований, требующей большой точности, необходимо применять набор фильтров с числом каналов минимум 256. Данный вариант обеспечивает погрешность вычисления частоты и интенсивности речевого сигнала не более чем 0,08 и 2,83 процента соответственно.

Изменение типа данных дает лишь незначительное уменьшение погрешности (изменения влияют на шестую значимую цифру числа), поэтому в целях экономии оперативной и постоянной памяти компьютера рекомендуется использовать наименьший по размеру тип данных – вещественный размером 4 байта. Этот же тип данных обеспечивает наилучшее быстроедействие на всех применяемых в ходе исследований процессорах.

Для выбранного типа данных при 128 каналах фильтрации нормированное время выполнения алгоритма свертки составляет 5,71 секунд (2,86 секунд) в реализации без учета (с учетом) краевого эффекта. При 256 каналах фильтрации нормированное время составляет 17,14 секунд (11,43 секунд) в реализации без учета (с учетом) краевого эффекта.

Представленные выше оценки времени проводились на компьютере с процессором AMD Athlon XP 1800+. Из четырех рассмотренных комбинаций самая быстрая имеет дефицит времени почти в 6 раз. На сервере с двумя двухядерными процессорами AMD Opteron 2216 дефицит времени сокращается примерно до 3 раз. Для проведения исследований речевого сигнала в реальном масштабе времени планируется реализация алгоритмов обработки в виде многопоточного приложения и применение многопроцессорного и/или многоядерного компьютера.

### Литература

1. Фланаган Дж. Анализ, синтез и восприятие речи / Дж. Фланаган / пер. с англ. / под ред. А.А. Пирогова. – М. : Связь, 1968. – 396 с.
2. Бондаренко В.П. Выделение особенностей структуры речевого сигнала / В.П. Бондаренко, В.П. Коцубинский, Р.В. Мещеряков // Сборник трудов XII сессии Российского акустического общества. – Т. 3. – М., 2003. – С. 63–66.
3. Конев А.А. Сегментация речевого сигнала / А.А. Конев, А.А. Пономарев // Сборник трудов XVI сессии Российского акустического общества. – Том III. – М. : ГЕОС, 2005. – С. 44–47.
4. Конев А.А. Программный комплекс для исследования речи / А.А. Конев, Е.Ю. Костюченко, А.А. Пономарев // Сборник трудов XVII сессии Российского акустического общества. – Т.3. – М. : ГЕОС, 2006. – С. 23–27.
5. Щеголев И.Е. Структура программного комплекса для реабилитации онкологических больных // Научная сессия ТУСУР – 2008 : Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых. – Ч. 3. – Томск : В-Спектр, 2008. – С. 197–200.

6. ГОСТ Р 50840-95. «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости». – М. : Госстандарт России, 1996.

---

**Щеголев Илья Евгеньевич**

Аспирант кафедры комплексной информационной безопасности  
электронно-вычислительных систем ТУСУРа

Тел.: (3822) 413-426

Эл. почта: sie@keva.tusur.ru

I.E. Shegolev

**The performance time estimation and speech signal processing error at use man's acoustical model**

In article are considered results of experiments on testing of acoustical model on various software and hardware. Recommendations about use of data various types and processor architecture are resulted.

---