УДК 612.391:004.934

МЕТОД ИЗМЕРЕНИЯ ЧАСТОТЫ ОСНОВНОГО ТОНА С МЕЖПЕРИОДНЫМ НАКОПЛЕНИЕМ РЕЧЕВОГО СИГНАЛА

Савченко В.В., *д.т.н., профессор, профессор кафедры «Математика и информатика» Нижегородского государственного лингвистического университета; e-mail: svv@lunn.ru.*

MEASUREMENT METHOD OF PITCH WITH INTER PERIOD ACCUMULATION OF SPEECH SIGNAL

Savchenko V.V.

A new method for measuring the pitch frequency of a speech signal in conditions of increased noise is proposed. The effectiveness of the method is achieved due to the effect of interperiodic accumulation. This effect is realized in a multichannel frequency measurement system using several parallel connected signal recirculator with adjustable delay periods in the feedback circuits. The effectiveness of the method has been studied theoretically and experimentally. The accuracy and sensitivity of the method are estimated as a function of the interference intensity at the input of the measurer. It is shown that for a signal-to-interference ratio of 20 dB or more, the error of the new method does not exceed (1...2) % of the nominal value of the frequency. It is agrees well with the potentially achievable accuracy under the conditions in question. In conditions of high noise the gain the threshold value of the signal-to-interference ratio in comparison with the world analogs is 4-5 dB or more.

Key words: speech, speech signal, pitch frequency, automatic speech processing, speech technology.

Ключевые слова: речь, речевой сигнал, частота основного тона, автоматическая обработка речи, речевые технологии.

Предложен новый метод измерения частоты основного тона речевого сигнала в условиях повышенного зашумления. Результативность метода достигается за счет эффекта межпериодного накопления. Указанный эффект реализуется в системе многоканального измерения частоты с использованием нескольких параллельно включенных накопителей-рециркуляторов с регулируемыми периодами задержки сигнала в цепях обратной связи. Эффективность метода исследована теоретически и экспериментально. По результатам проведенного исследования даны оценки точности и чувствительности метода в зависимости от интенсивности помехи на входе измерителя. Показано, что при отношении сигнал-помеха 20 дБ и более погрешность нового метода не превышает (1…2) % от номинала частоты основного тона, что хорошо согласуется с потенциально достижимой точностью в рассматриваемых условиях. В условиях повышенного зашумления выигрыш в пороговой величине отношения сигнал-помеха по сравнению с мировыми аналогами составляет 4-5 дБ и более.

Введение

Частота основного тона (ЧОТ) относится к наиболее информативным акустическим характеристикам речевого сигнала [1-3] и в этом качестве играет важную роль при автоматической обработке речи в системах самого разного назначения [4, 5], в том числе на выходе телефонного канала связи [6]. Исследования в данной области непрерывно продолжаются [7, 8], в том числе, по актуальному направлению [9, 10] повышения помехоустойчивости измерителя ЧОТ. Однако из-за известной сложности такого рода задач в теории до сих пор отсутствует ее эффективное решение. Так, речевой сигнал принципиально вариативен по своей тонкой структуре [11], модулирован по амплитуде [12] и не стабилен по динамике [13, 14]. При этом действие акустического (фонового) шума [15, 16] только усиливает все перечисленные факторы. Как следствие, большинство известных методов измерения ЧОТ хорошо зарекомендовало себя в «тепличных», бесшумных условиях и, вместе с тем, сильно теряет по эффективности при действии случайных помех средней мощностью минус 10 дБ и выше относительно мощности сигнала [17]. А это, между прочим, распространенные условия [3] производства и восприятия речи в процессе коммуникаций. Проблема обостряется условиями малых выборок наблюдений в пределах интервалов вокализации речевого сигнала конечной длительности T [18]. Разработке и исследованию метода измерения ЧОТ повышенной помехоустойчивости посвящена настоящая статья. В ней используется математический аппарат теории сигналов [19, 20].

Постановка задачи

Согласно акустической теории речеобразования [2], сигнал $x(t)$ на выходе речевого тракта диктора на интервалах действия гласных фонем имеет линейчатый частотный спектр. Его минимальная (нижняя) частота F_0 и определяет текущее значение ЧОТ. В диапазоне ее значений (80…160) Гц, характерном для разговорной речи мужчин, при T=150…200 мс [11] будем иметь порядка $M = TF_0 =$ 15…30 периодов основного тона в пределах вокализованных отрезков речевого сигнала. А это серьезный стимул для применения межпериодного накопления [21] в качестве радикального средства повышения помехоустойчивости обработки.

Известно [19-21], что указанное накопление в общем случае реализуется по схеме гребенчатого фильтра

накопления (ГФН), настроенного на частоту F_0. В условиях априорной неопределенности теория [20] рекомендует многоканальную систему обработки сигнала с использованием набора из N параллельно включенных ГФН, перекрывающих своими частотными характеристиками весь анализируемый диапазон. Ее структурная схема изображена на рис. 1, где приняты следующие обозначения: АД – амплитудный (квадратичный) детектор, ФНЧ – фильтр нижних частот. Система АД–ФНЧ служит для измерения средней мощности или дисперсии речевого сигнала $y_i(t)$ на выходе ГФН в составе i-го канала. Решение в отношении оптимальной оценки частоты $\widehat{F}_0$ здесь принимается по принципу максимума средней мощности $P(y_i)$ накопленного сигнала $y_i(t)$. Вопросы практического осуществления и анализа многоканальной системы (рис. 1) в задаче измерения ЧОТ подробно рассматриваются далее.

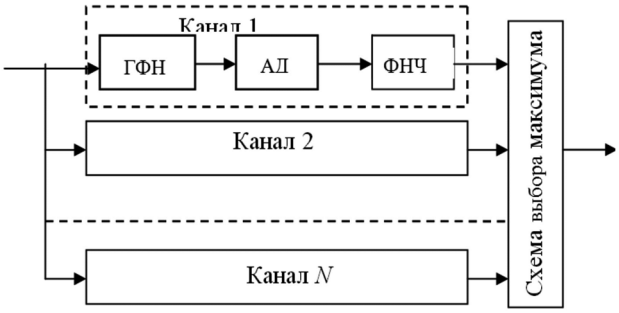

Рис. 1. Система оптимального измерения частоты

Синтез алгоритма

При большом числе каналов N>>1 (рис. 1) разработчики отдают предпочтение линейным рекурсивным фильтрам простейшего типа – по схеме рециркулятора [21]. Его динамика описывается рекуррентным уравнением первого порядка

$$y_i(t) = x(t) + by(t - T_i), \quad (1)$$

где $T_i = 1/F_i -$ период накопления сигнала $x(t)$, $b < 1 -$ коэффициент его усиления в цепи обратной связи. Чем ближе значение b к единице, тем более выражен в (1) эффект накопления. Однако на практике следует учитывать конечную длительность речевого сигнала на интервалах его вокализации. Поэтому существует оптимальное значение b_0 параметра b из условия максимизации достигаемого выигрыша в отношении сигнал–помеха (ОСП) за счет накопления сигнала. В первом приближении можно записать (см., напр., монографию [21], ф. 5.4.9 на с. 202) $b_0 \approx 1 - 1,27/M$, или примерно 0,95 в рассматриваемой нами задаче. При этом период накопления сигнала в каждом канале (рис. 1) соответствует определенному варианту ЧОТ $F_i = 1/T_i$ в пределах диапазона ее ожидаемых значений $[F_{\min}; F_{\max}]$. В режиме дискретного времени t = 0, ΔT, $2\Delta T$, … с частотой дискретизации речевого сигнала $F = const$ будем иметь $T_i = T_{i-1} + \Delta T$, где $\Delta T = 1/F$ – период взятия отсчетов. При равенстве $T_0 = 1/F_{\max} - \Delta T$ получаем

$$T_i = i\Delta T, i = \overline{1, N}, \quad (2)$$

где число каналов $N = (T_N - T_1)/\Delta T = F(F_{\min}^{-1} - F_{\max}^{-1})$. Например, при частоте дискретизации сигнала F = 8 кГц, согласованной с полосой пропускания стандартного телефонного канала связи [3], в расчете на мужские голоса будем иметь $N = 8000(80^{-1} - 160^{-1}) = 8000/160 = 50$.

Дополняя (1) оценкой средней мощности накопленного сигнала по формуле эмпирической дисперсии [19]

$$P(y_i) = (\tau F)^{-1} \sum_{k=1}^{\tau F} y_i^2(t + k\Delta T), \quad (3)$$

где τ – длительность речевого фрейма или интервала квазистационарности речевого сигнала, предполагаемого центрированным, получим квазиоптимальный алгоритм измерения ЧОТ на основе межпериодного накопления:

$$\widehat{F}_0 = 1/T_m, m = \operatorname{Arg}\max_{i \le N} P(y_i). \quad (4)$$

Его эффективность может быть охарактеризована как точностью результирующей оценки ЧОТ, так и помехоустойчивостью или надежностью обработки речевого сигнала в условиях действия случайных помех [20].

Анализ точности

Точность измерения ЧОТ согласно алгоритму (1)-(4) определяется, главным образом, его инструментальной погрешностью. При учете достаточно высокой разрешающей способности ГФН по частоте [21] $\Delta f \approx \pi^{-1} F_i \times$ $\times (1 - b_0)/(1 + b_0)$, которая при $F_i \le 160$ Гц и b_0 = 0,95 не превышает 1,3 Гц, определим инструментальную погрешность алгоритма через величину частотной расстройки

$$\Delta F_i = 0,5(1/T_i - 1/T_{i+1}) =$$
$$= 0,5\Delta T/(T_i T_{i+1}) = 0,5/(FT_i T_{i+1}).$$

двух соседних каналов измерителя ЧОТ (рис. 1). Нетрудно понять, что она ограничена сверху предельным уровнем $0,5/(FT_1^2) = 0,5F_{\max}^2/F$. Например, в рассматриваемом диапазоне частот будем иметь $\Delta F_i \le 0,5 \times$ $\times 160^2/8000 = 1,6$ Гц. Отметим, что полученный результат отвечает требованиям госстандарта [22] к частотным измерителям резонансного типа и при этом превышает на (30…40) % показатели точности современных методов измерения ЧОТ из обзорной работы [17]. Аналогичный вывод можно сделать и в отношении помехоустойчивости предложенного метода.

Со ссылкой на монографию [21] при действии гауссовских помех помехоустойчивость метода может быть охарактеризована выигрышем ОСП по мощности q^2 величиной порядка $W \approx 0,8M$. В нашей задаче она составляет примерно 12 дБ. Сказанное в равной мере относится как к помехам типа белого шума, так и к подобным речевому сигналу помехам с «окрашенными» спектрами мощности [15]. Второй тип помех имеет неотличимую от полезного сигнала внутрипериодную тонкую (формантную) структуру [18] и поэтому представляет неразрешимую на данный момент проблему [17] для большинства известных методов измерения ЧОТ. Одна-

ко, с точки зрения межпериодной обработки (1) подобные помехи неотличимы от белого шума, т.к. не содержат в своем спектре достаточно мощных периодических компонент в пределах рабочего диапазона частот. Поэтому вне зависимости от типа действующих помех эффективность предложенного метода в статистическом смысле близка к потенциально достижимой эффективности измерения частоты. В пересчете к среднеквадратичной величине ошибки измерений [20] через эффективную длительность речевого сигнала, примерно равную WT_i, при q^2 =20 дБ получим $\sigma_0^* \approx 1 \div \div\sqrt{q^2 (WT_i)^2} \le F_{\max} / (W\sqrt{q^2}) = 160 / (0{,}8 \times 20 \times 10) = 1$ Гц. Подчеркнем, это справедливо при почти идеальных условиях. Но, например, уже при q^2 =10 дБ будем иметь как минимум $\sigma_0^* = 3{,}16$ Гц, а при q^2 =0 дБ – в три раза больше: $\sigma_0^* = 10$ Гц. Исследование эффективности предложенного в статье метода в условиях существенного зашумления речевого сигнала проводится далее экспериментальным путем. Об актуальности этой задачи свидетельствует, например, следующий факт: заявленные в работе [9] в качестве среднеквадратичной величины ошибки измерения ЧОТ (0,039…0,045) Гц при применении так называемого «*метода активного восприятия*»[1] для случая действия белого гауссовского шума при двух рассмотренных выше значениях ОСП более чем на два порядка (!) превышают потенциальную точность оптимального измерителя частоты.

Программа и методика эксперимента

Исследования проводились в два этапа. На первом этапе условный диктор (автор статьи) несколько раз прочитал через микрофон художественный текст из первой главы поэмы А.С. Пушкина «Евгений Онегин». И каждый раз записал прочитанный текст на персональный компьютер в виде звукового файла. Затем автоматически из каждой записи были выделены в отдельные файлы все ее вокализованные отрезки. При этом использовалась известная [23] методика. При длительности каждой записи 3 мин. и более в результате было получено не менее 200 файлов речевого сигнала $x(t)$ длительностью 150-200 мс каждый. Согласно рекомендациям работы [24] этого вполне достаточно для получения в дальнейшем надежных статистических оценок. Тем самым на первом этапе экспериментальных исследований был сформирован необходимый речевой материал.

На втором этапе полученные файлы были подвергнуты компьютерной обработке согласно разработанному выше алгоритму (1)-(4). Для этого была создана и использована специальная авторская программа «*Speech Transform*», включающая в себя рециркулятор с регулируемым периодом задержки T_i сигнала $x(t)$ в цепи обратной связи. Главное окно программы показано на рис. 2. Здесь в правой части указаны параметры алгоритма и введено обозначение $n = T_i F$ для относительной величины периода T_i. Оптимальная оценка ЧОТ определяется в данной программе через частоту дискретизации речевого сигнала простым соотношением вида $\hat{F}_0 = 1 / T_m = F / n^*$, где n^* – значение параметра для канала (рис. 1) с максимальным сигналом на выходе. В нашем случае мы получили n^* = 81 и, следовательно, $\hat{F}_0 = 8000 / 81 = 98{,}8$ Гц.

В верхней части окна показана временная диаграмма одного из файлов $x(t)$. В данном случае это был сигнал ударной гласной «А» из слова «*прАвил*». В нижней части окна показана временная диаграмма сигнала на выходе ГФН в составе определенного канала измерителя

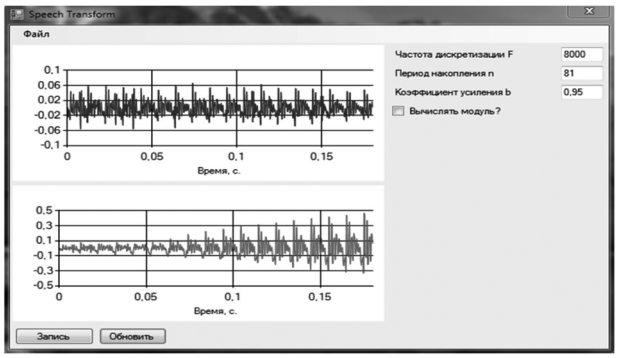

Рис. 2. Скриншот главного окна компьютерной программы

ЧОТ (рис. 1). В нашем случае это был оптимальный, *m*-й канал с сигналом $y_m(t)$ максимальной мощности. Из сопоставления двух диаграмм на рис. 2 хорошо виден эффект межпериодного накопления речевого сигнала в ГФН. Его исследованию и был посвящен, главный образом, весь дальнейший эксперимент. При этом использовалась аддитивная модель смеси сигнала с помехой $\tilde{x}(t) = x(t) + \eta(t)$, в рамках которой помеха формировалась программным способом с использованием стандартного генератора белого гауссовского шума. Его дисперсия, а вслед за ней и ОСП варьировались в эксперименте в широких пределах. И каждый раз вычислялась средняя мощность $P(y_m)$ накопленного сигнала $y_m(t)$ на интервале в один речевой фрейм длительностью τ =10 мс или на интервале в $\tau F = 80$ отсчетов данных. Полученные результаты представлены на рисунках ниже.

Основные результаты и выводы

Для оценки чувствительности алгоритма (1)-(4) в качестве важнейшей технической характеристики измерителей частоты [22] на рис. 3 представлена зависимость избирательной способности метода по частоте $\gamma_m(i) = P(y_{m+i}) / P(y_m)$ через относительную величину средней мощности накопленного сигнала $P(y_{m+i})$ в каждом отдельном канале системы обработки (рис. 1), начиная с оптимального, *m*-го канала. При ее вычисле-

[1] А может ли быть восприятие не активным? На этот риторический вопрос в свое время четко ответил авторитетный ученый-психолог Дж. Гибсон: «Восприятие – это активный процесс извлечения информации об окружающем мире, включающий в себя реальные действия по обследованию того, что воспринимается» (см. его монографию «Экологический подход к зрительному восприятию» - М.: Прогресс, 1988).

ниях использовалась выборка из 200 независимых реализаций вокализованных отрезков речевого сигнала $x(t)$, а ОСП составляло 20 дБ.

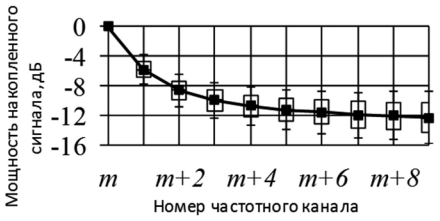

Рис. 3. Избирательная способность по частоте

Как видим, зависимость в данном случае носит ярко выраженный спадающий характер, причем, особенно сильно – в начале координат. Ее первое значение $\gamma_m(1)$ составляет порядка минус 6 дБ. В этих условиях чувствительность предложенного метода может быть охарактеризована половиной частотного интервала между двумя соседними каналами измерителя ЧОТ (рис. 1). Путем несложных вычислений получаем $\Delta F_m = 0,5 \div \div (FT_mT_{m+1}) \approx 0,6$ Гц. Это с запасом удовлетворяет требованиям действующего стандарта [22].

Для оценки помехоустойчивости метода далее была исследована зависимость его избирательной способности $\gamma^* \overset{\Delta}{=} \gamma_m(1)$ от ОСП. Для этого было сформировано достаточно представительное множество независимых реализаций белого гауссовского шума $\eta(t)$ разной дисперсии в расчете на фиксированную среднюю мощность речевого сигнала $x(t)$. Полученная зависимость $\gamma^*(q^2)$ представлена в виде диаграммы на рис. 4.

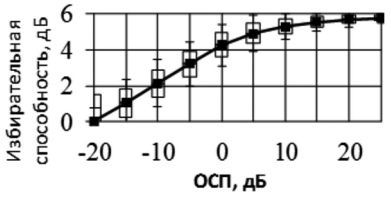

Рис. 4. Зависимость избирательной способности от ОСП

Пороговая величина ОСП здесь составила (-4… -10) дБ. Это на 4-5 дБ лучше, чем в методе, предложенном в работе [14], который был разработан автором специально для использования в условиях повышенного зашумления.

В развитие полученных результатов далее была исследована зависимость $\gamma^*(q^2)$ при действии случайной помехи с окрашенным спектром мощности. В эксперименте она была сформирована пропусканием нормального белого шума через ФНЧ с эффективной полосой пропускания, равной $0,1F$ = 800 Гц. При этом для каждой точки q^2 использовалась выборка из L = 260 независимых файлов окрашенной помехи $\eta(t)$. Полученные оценки, как и ожидалось, по форме повторили зависимость, представленную на рис. 4. Потери в ОСП при этом не вышли за пределы (1…2) дБ – на уровне статистической погрешности измерений. В самом деле, следуя гауссовской аппроксимации усредненной выборочной величины γ^*, воспользуемся в качестве ее вероятностной характеристики классическим выражением для половины длины доверительного интервала в его относительном выражении $\delta = z_p / \sqrt{L}$ [24]. Здесь z_p – коэффициент надежности или «доверия» на уровне значимости α, определяемый корнем уравнения $\Phi(z_p) = p$ с интегралом вероятности в левой части. При равенстве $\alpha = 0,1$ (соответствующая доверительная вероятность равна $p = 1 - \alpha$ = 0,9) и объеме выборки L = 260 по таблицам нормального распределения [25] находим $z_{0,9} \approx 1,65$. И, следовательно, получаем $\delta \approx 10\%$. При учете номинального значения характеристики $\gamma^* \approx 6$ дБ (см. рис.4) данный результат представляется вполне приемлемым с точки зрения точности проведенного исследования.

Заключение

В статье предложен новый метод измерения ЧОТ на основе межпериодного накопления речевого сигнала, исследована его эффективность. Для реализации эффекта накопления используется ГФН простейшей структуры – рециркулятор с регулируемым периодом задержки сигнала в цепи обратной связи. Благодаря накоплению предложенный метод обеспечивает выигрыш по сравнению с его известными аналогами как в точности измерения ЧОТ, так и в его чувствительности.

Полученные результаты и сделанные по ним выводы позволяют рекомендовать данный метод для практического применения в системах автоматической обработки речи при их работе в условиях повышенного зашумления.

Литература

1. Christensen M, Jakobsson A. Multi-pitch Estimation. – Morgan and Claypool, 2009. – 432 p.

2. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 304 с.

3. Савченко В.В. Оценка качества цифровой передачи речи по конечной выборке речевого сигнала // Электросвязь. 2017. № 3. С. 52-57.

4. Лебедева Н.Н., Каримова Е.Д., Казимирова Е.А. Анализ речевого сигнала в исследованиях функционального состояния человека // Биомедицинская радиоэлектроника. 2015. № 2. С. 3-12.

5. Андреева Н.Г., Смирнова Т.А. Восприятие синтезированных моделей одноформантных гласных с разной частотой основного тона // Сенсорные системы. 2014. Т. 28. № 4. С. 13-21.

6. Чернобельский С.И. Сравнение результатов акустического анализа голоса при различных способах его записи // Вестник оториноларингологии. 2014. № 1. С. 41-43.

7. Алимурадов А.К. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов // Труды Московского физико-технического института. 2015. Т. 7. № 3 (27). С. 56-68.

8. Вишнякова О.А., Лавров Д.Н. Гибридный алгоритм выделения частоты основного тона // Математические структуры и моделирование. 2016. № 1 (37). С. 59-65.

9. Гай В.Е. Метод оценки частоты основного тона в условиях помех // Цифровая обработка сигналов. 2013. № 4. С. 65-71.

10. Архипов И.О., Гиниятуллин Р.Р. Автокорреляционный выделитель основного тона с предварительным оцениванием частоты колебаний голосовых связок // В сборнике: «Молодые ученые – ускорению научно-технического прогресса в XXI веке». – Ижевск: Изд-во: ИННОВА, 2016. С. 421-428.

11. Savchenko V.V., Savchenko A.V. Information Theoretic Analysis of Efficiency of the Phonetic Encoding–Decoding Method in Automatic Speech Recognition // Journal of Communications Technology and Electronics. 2016. Vol. 61. No. 4. P. 430-435.

12. Азаров И.С., Вашкевич М.И., Петровский А. Алгоритм оценки мгновенной частоты основного тона речевого сигнала // Цифровая обработка сигналов. 2012. № 4. С. 49-57.

13. Вольф Д.А., Мещеряков Р.В. Модель и программная реализация сингулярного оценивания частоты основного тона речевого сигнала // Труды СПИИРАН. 2015. № 6. С. 191-209.

14. Савченко В.В. Тестирование случайных временных рядов на стационарность на основе принципа минимума информационного рассогласования // Известия вузов. Радиофизика. 2017. Т. 60. № 1. С. 89-96.

15. Savchenko V.V. Enhancement of the Noise Immunity of a Voice-Activated Robotics Control System Based on Phonetic Word Decoding Method // Journal of Communications Technology and Electronics. 2016. Vol. 61. No. 12. P. 1374 -1379.

16. Савченко В.В., Акатьев Д.Ю., Афонин М.В. Автоматическое распознавание речи на фоне шума // Современные тенденции развития науки и технологий. 2015. № 6-2. С. 99-102.

17. Hasan M.A. Pitch Detection Algorithm Based on Windowless Autocorrelation Function and Modified Cepstrum Method in Noisy Environments // IJCSNS International Journal of Computer Science and Network Security. 2017. Vol.17. No.2. P. 106-112.

18. Savchenko V.V. The Principle of the Information-Divergence Minimum in the Problem of Spectral Analysis of the Random Time Series Under the Condition of Small Observation Samples // Radiophysics and Quantum Electronics. 2015. Vol. 58. No. 5. P. 373-379.

19. Радиотехнические системы / Под ред. Ю.М. Казаринова. – М.: Издательский центр «Академия», 2008. – 592 с.

20. Лезин Ю.С. Введение в теорию и технику радиотехнических систем. – М.: Радио и связь, 1986. – 140 с.

21. Лезин Ю.С. Оптимальные фильтры и накопители импульсных сигналов. - М.: Советское радио, 1969. - 448 с.

22. ГОСТ 12692-67. Измерители частоты резонансные. Методы и средства поверок. - М.: Изд-во стандартов, 1967. - 7 с.

23. Savchenko V.V. Ponomarev D.A. Automatic Segmentation of Stochastic Time Series Using a Whitening Filter // Optoelectronics, Instrumentation and Data Processing. 2009. Vol. 45. No. 1. P. 37-42.

24. Савченко В.В. Определение объема контрольной выборки в условиях априорной неопределенности по принципу гарантированного результата // Научные ведомости БелГУ. Серия: Экономика. Информатика. 2015. № 1 (198). Вып. 33/1. С.74-78.

25. Большаков В.Д. Теория ошибок наблюдений. - М.: Недра, 1983. - 223 с.