УДК 681.327

д.т.н., проф. Хорошко В.А, Демьян Н.И.,
Осмаловский В.А. (ГУИКТ)

ОЦЕНКИ ПАРАМЕТРОВ МОДЕЛЕЙ РЕЧЕВЫХ СИГНАЛОВ

Введение

Под понятием звука подразумевается механические колебания упругой среды (воздуха, воды, металл т.д.), субъективно воспринимаемые органами слуха человека в диапазоне частот от 16 до 20000 Гц.

Источником образования речевого или звукового сигнала являются вибрирующие, колеблющиеся тела и механизмы, такие как голосовые связки человека, движущие элементы машин, телефонные аппараты, звукоусилительные системы и т.д.

Поскольку с помочью речи осуществляется передача информации, возможны, по меньшей мере, два подхода к количественному описанию этого процесса. Один из них основан на теории информации [1], в соответствии с которой речь можно охарактеризовать ее информативным содержанием. Другой способ описания речи заключается в представлении ее в виде сигнала, т.е. акустического колебания. Хотя идеи теории информации играют важную роль при построении сложных систем связи, наиболее полезными на практике все же оказывается представление реи в виде колебания, порожденного речевым трактом человека.

Основная часть

Математическое ожидание речевых сигналов, как правило, содержит неизвестные параметры. Их оценивание – важная задача обработки. Обычно основными параметрами речевых сигналов считается ферментные частоты, период (частота) основного тона, различные коэффициенты дробно-рациональной передаточной функции линейной модели голосового тракта (коэффициенты отражения, дисперсия, среднее число пересечений нулевого уровня и т.п.)

Поэтому для оценки параметров речевых сигналов широко используются модели человеческого тракта (акустическая, электрическая и математическая).

Упрощенной акустической моделью системы резонаторов голосового тракта является сочленение коротких цилиндрических труб. Представленная на рис.1 модель состоит из трех основных секций и четвертой, дополнительной, для имитации губ [1]. Влияние ротовой полости при этом не учитывается, то для большинства вокализованных звуков вполне приемлемо.

Первый резонатор с помощью поперечного сечения A_1 и длиной l_1 имитирует гортань и заднюю ротовую полость (до языка), второй (A_2 и l_2) – участок сужения между языком и твердом небам, третий (A_3 и l_3) – переднюю ротовую полость, четвертый (A_4 и l_4) – проход между губами.

Эта акустическая модель системы резонаторов является достаточно грубой [1]. Усложняя ее путем добавления параллельных каналов, имитирующих носовую полость, можно получить более точную акустическую модель голосового тракта человека.

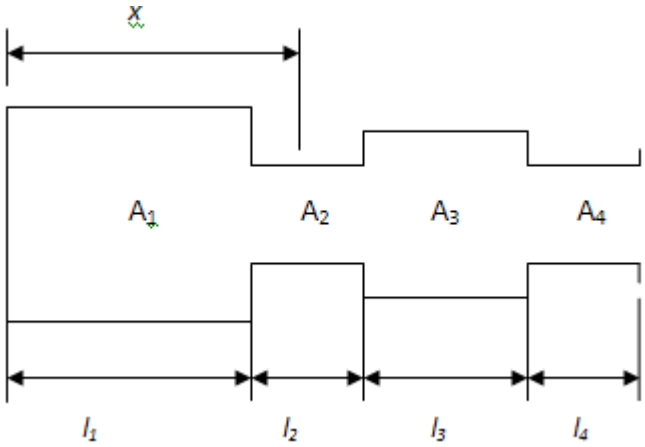

Рис.1. Акустическая модель системы резонаторов голосового тракта

Для перехода от акустической модели к электрической используется метод электроакустических аналогий [2], в соответствии с которым отрезок трубы длиной l с неизменным поперечным сечением A представляют электрической схемой (рис.2).

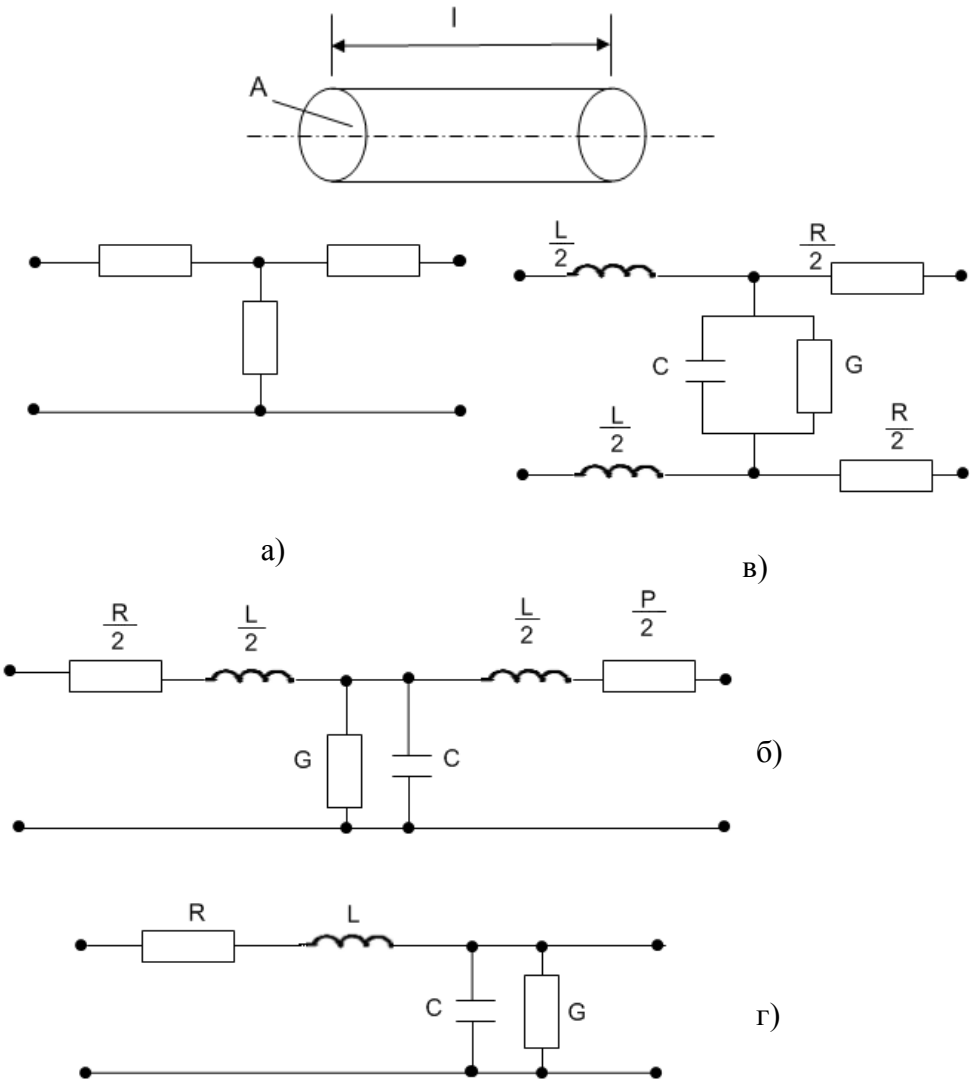

Рис.2. Электрические эквиваленты отрезка трубы

В предположении, что труба гладкая и имеет жесткие стенки, параметры электрических эквивалентов равны:

$$L=\frac{\rho l}{A};C=\frac{Al}{\rho c^2};R=\frac{sl}{A^2}\sqrt{\frac{\omega\rho\mu}{2}};G=sl\frac{\eta-1}{vc^2}\sqrt{\frac{K_h\omega}{2C_p\rho}}; \quad (1)$$

где ρ - плотность воздуха; c -скорость звука; l - длина трубы; s - длина окружности трубы; μ - коэффициент вязкости; K_h - коэффициент теплопроводности; η - адиабатическая постоянная; C_p - удельная теплоемкость воздуха при постоянном давлении.

При этом в методе электроакустических аналогий [2] звуковое давление рассматривают как эквивалент напряжения, а скорость воздушного потока – как эквивалент тока. Инертность воздушной массы аналогична индуктивности L, а упругость сжимаемого объекта воздуха – емкости C. Сопротивление R представляет потери, обусловленные вязким трением о стенки требы. Эквивалентная проводимость G описывает потери, вызванные теплопроводностью стенок трубы.

В теории сигналов и систем существуют собственные традиции моделирования, существенно использующие такие понятия как «спектр сигналов», «импульсная характеристика», «частотная характеристика системы». Для аналогического описания таких моделей используется небольшое количество базовых соотношений, и в первую очередь – соотношение свертки и преобразования Фурье.

Так, при цифровой обработке сигналов, отклик y_n каждого из полосовых фильтров в декодере полосного вокодера может быть вычислен в соответствии с одним из следующих соотношений:

$$y_n=\sum_{k=0}^{N_{fft}-1}h_k x_{n-k}\ , \quad (2)$$

$$y_n=\sum_{r=0}^{N_{fft}-1}H_r X_r \exp\left(j\frac{2\pi}{N_{fft}}rn\right), \quad (3)$$

где h_k и H_r - выборки импульсной и частотной характеристик полосового цифрового фильтра, соответственно; x_n и X_r - выборки возбуждаемого сигнала и его дискретного Фурье-образа, соответственно; N_{fft} -параметр алгоритма быстрого преобразования Фурье (БПФ).

Соотношения (2) и (3) предназначены для вычисления с применение алгоритма БПФ. В качестве альтернативы этим соотношениям, можно использовать уравнение цифровой фильтрации:

$$y_n=\sum_{k=0}^{N}a_k x_{n-k}-\sum_{k=1}^{N}b_k y_{n-k}\ , \quad (4)$$

где a_k и b_k - коэффициенты полосового цифрового фильтра.

Точность оценивания некоторого параметра λ допустимо характеризовать как среднее значение квадрата ошибки (СКО)

$$\varepsilon_\lambda^2=E\left(\widehat{\lambda_n}-\lambda\right)^2, \quad (5)$$

где $\widehat{\lambda_n}$ - оценка параметров по выборке объема n; E - усреднение с функцией плотности ФПВ $\widehat{\lambda_n}$ и λ, если λ - случайная величина. Когда оценка неизвестного детерминированного λ перемещения, т.е. $E\hat{\lambda}_n=\lambda$, то ε_λ^2 является дисперсией $D\hat{\lambda}_n$ оценки.

При поступлении оценки всегда желательно, чтобы ε_λ^2 была по возможности минимальной. Это обычное в математической статистике стремление при обработке речевых сигналов иметь дополнительный аспект: «хорошим» моделям сигнала соответствуют «хорошие» оценки их параметров. Поэтому, сравнивая значения минимально возможных СКО оценивания параметров различных моделей можно судить и о возможностях применения самих моделей.

При рассмотрении параметров моделей будем осуществлять оценку нижних границ СКО, ограничившись при этом несмещенными оценками. Оценки параметров строятся по наблюдаемой реализации $x_1, x_2, ..., x_n$ речевого сигнала (выборке объема n), условии ФПВ которого $p\left(x_1, x_2, ..., x_n / \lambda\right) = p\left(x_n^1 / \lambda\right)$ известна: если выполняется некоторые условия регулярности, то дисперсия оценки удовлетворяет неравенству Крамера-Рао

$$D\hat{\lambda}_n \geq \left\{ E\left[\frac{\partial}{\partial \lambda} \ln p\left(x_n^1 / \lambda\right) \right]^2 \right\}^{-1}. \tag{6}$$

Усреднение (6) выполняется по $p\left(x_n^1 / \lambda\right)$. Оценку, имеющую нижнюю граничную дисперсию $D\hat{\lambda}_{эф}$ принято называть эффективной. Эффективностью данной оценки параметра называют отношение дисперсии эффективной оценки $D\hat{\lambda}_{эф}$, к дисперсии $D\hat{\lambda}_n$ данной оценки. Заметим, что эффективная оценка параметра существует не всегда. Если представляет интерес не оценка $\hat{\lambda}_n$ параметра λ, а оценка $\hat{g}_n$ дифференцируемой функции параметра $g\left(x\right)$, то

$$D\hat{g}_n \geq \left[\frac{dg\left(\lambda\right)}{d\lambda} \right]^2 D\hat{\lambda}_n. \tag{7}$$

Пусть оцениванию подлежит не один параметр, несколько, образующих вектор $\left(\lambda_1, \lambda_2, ..., \lambda_n\right)^T = \lambda$, параметров. Тогда корреляционная матрица ошибок $R_{\hat{\lambda}}$ удовлетворяет следующему неравенству Крамера-Рао:

$$R_{\hat{\lambda}} \geq J^{-1}, \tag{8}$$

где элементы $R_{\hat{\lambda}}$

$$R_{ij} = E\left(\hat{\lambda}_i - \lambda_i\right)\left(\hat{\lambda}_j - \lambda_j\right), i = 1, 2, ..., m, \tag{9}$$

J - информационная матрица Фишера с элементами

$$J_{ij} = E\left\{ \frac{\partial \ln p\left(x_n^1 / \lambda\right)}{\partial \lambda_i} \cdot \frac{\partial \ln p\left(x_n^1 / \lambda\right)}{\partial \lambda_j} \right\}, \tag{10}$$

J^{-1} - матрица обратная J.

Если в выражении (4) достигается равенство, то оценки называются совместно эффективными.

Рассчитаем минимальные дисперсии оценок параметров моделей по выражению из [4] при $m = 1$ и получаем

$$\Im_2\left(t\right) = r_1 \xi_2\left(t-1\right) + \Im_1\left(t\right)$$

или

$$x_t = r_1 x_{t-1} + b\xi_t, \tag{11}$$

где $E\xi_t = 0, E = \xi_t^2 = 1, \xi_t$ - стационарная последовательность некоррелированных гауссовских случайных величин: $b \prec \infty, Ex_t^2 \prec \infty$, т.е. $|r_1| \prec 1$. Модель (II) является простейшим описаним речового сигнала на участках произнесения невокализованных звуков. Пусть $x_0 = const, t = 1, 2, \ldots$. Тогда

$$p(x_1, x_2, \ldots, x_n / r_1) = \frac{1}{b^n}\prod_{i=1}^{n} p\left[\xi_i = \frac{1}{b}(x_i - r_1 x_{i-1})\right] = \frac{1}{b^n}\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2b^2}\sum_{i=1}^{n}(x_i - r_1 x_{i-1})^2\right\}, \tag{12}$$

где $\frac{1}{b^n}$ - значение якобиана перехода от переменных $x_1, x_2, \ldots, x_n$ к $\xi_1, \xi_2, \ldots, \xi_n$, а $p(\xi_i)$ - гауссовская ФПВ случайной величины ξ_i. Производная $\frac{\partial}{\partial r_1}\ln p\left(x_n^1 / r_1\right) = \frac{1}{p\left(x_n^1 / r_1\right)}\frac{\partial}{\partial r_1} p\left(\frac{x_n^1}{r_1}\right)$,

где

$$\frac{\partial}{\partial r_1}\ln p\left(x_n^1 / r_1\right) = \frac{1}{b^n}\left(\frac{1}{\sqrt{2\pi}}\right)^n \sum_{i=1}^{n} x_{i-1}(x_i - r_1 x_{i-1}) e^{-\frac{1}{2b^2}\sum_{i=1}^{n}(x_i - r_1 x_{i-1})^2} =$$

$$= \frac{1}{b^2 \sum_{i=1}^{n} x_{i-1}(x_i - r_1 x_{i-1})\, p(x_1, x_2, \ldots, x_n / r_1)}. \tag{13}$$

Следовательно

$$\frac{\partial}{\partial r_1}\ln p\left(x_n^1 / r_1\right) = \frac{1}{b^2}\sum_{i=1}^{n} x_{i-1}(x_i - r_1 x_{i-1}) \tag{14}$$

и квадрат этого выражения можно записать так:

$$\left[\frac{\partial}{\partial r_1}\ln p\left(x_n^1 / r_1\right)\right]^2 = \frac{1}{b^4}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{i-1}x_{j-1}(x_i - r_1 x_{i-1})(x_j - r_1 x_{j-1}).$$

Теперь знаменатель формулы (6) равен выражению

$$\frac{1}{b^4}\sum_{i=1}^{n}\sum_{j=1}^{n} Ex_{i-1}x_{j-1} b\xi_i b\xi_j = \frac{1}{b^2}\sum_{i=1}^{n} Ex_{i-1}^2 \tag{15}$$

при получении, которого использованы соотношения

$$\xi_i = \frac{1}{b}(x_i - r_1 x_{i-1}), \xi_j = \frac{1}{b}(x_j - r_1 x_{j-1}), Ex_{i-1}x_{j-1}\xi_i\xi_j = \begin{cases} 0, i \neq j, \\ Ex_{i-1}^2, i = j. \end{cases}.$$

Последнее соотношение есть следствие независимости ξ_t от $x_1, x_2, \ldots, x_{t-1}$ в (II). Таким образом, из (15), (6) получаем дисперсию эффективной оценки

$$D\hat{r}_{1эф} = \frac{b^2}{\sum_{i=1}^{n} Ex_{i-1}^2}. \tag{16}$$

Выражение для Ex_{i-1}^2 можно получить из (II). Вычисляя дисперсию левой и правой частей уравнения, получаем

$$Ex_t^2 = r_1^2 Ex_{t-1}^2 + b, t = 1, 2, \ldots Ex_0 = x_0^2. \tag{17}$$

Интегрируя (17), приходим к соотношению:

$$Ex_t^2 = r_1^{2t} x_0^2 + b\sum_{i=0}^{t-1} r_1^{2t} = r_1^{2t} x_0^2 + \frac{b^2(1 - r_1^{2t})}{1 - r_1^2}$$

и

$$Ex_{i-1}^2 = r_1^{2(i-1)} x_0^2 + b^2 \frac{1-r_1^{2(i-1)}}{1-r_1^2}. \tag{18}$$

Поскольку x_t по условию – стационарная последовательность, переходным процессом в полученном соотношении следует пренебречь. Тогда $Ex_{i-1}^2 = \frac{b^2}{1-r_1^2}$ и

$$D\hat{r}_{1эф} = \frac{b^2}{nEx_t^2} = \frac{1}{n}\left(1-r_1^2\right). \tag{19}$$

Из (19) видно, что дисперсия эффективной оценки убывает с ростом n как $\frac{1}{n}$ и пропорциональная своеобразному отношению сигнал/шум: $\frac{b^2}{nEx_t^2}$. Роль шума выполняет порождающая последовательность. Чем ближе r_1 к границе устойчивости, тем меньше $D\hat{r}_{1эф}$. Так как ξ_t отражает турбулентные шумы в речеобразующей системе, то на невокализованных звуках всегда $\frac{b}{Ex_t^2}$ больше, чем на вокализованных [5]. Поэтому эффективная оценка $\hat{r}_1$ будет иметь меньшую дисперсию на участках произнесения вокализованных звуков.

Рассмотрим модель второго порядка , запись ее в форме уравнения авторегрессии

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \xi_t. \tag{20}$$

Теперь оценке подлежит вектор параметров $\left(a_1, a_2\right)^T = A$. Повторяя рассуждения, получаем

$$\begin{aligned} \frac{\partial}{\partial a_1} \ln p\left(X_n^1 / A\right) &= \frac{1}{b^2}\sum_{i=1}^{n}\left(x_i - a_1 x_{i-1} - a_2 x_{i-2}\right)x_{i-1}, \\ \frac{\partial}{\partial a_2} \ln p\left(X_n^1 / A\right) &= \frac{1}{b^2}\sum_{i=1}^{n}\left(x_i - a_1 x_{i-1} - a_2 x_{i-2}\right)x_{i-2}. \end{aligned} \tag{21}$$

Информационная матрица Финира [6]:

$$J = \frac{Ex_t^2}{b^2} n \begin{bmatrix} 1 & R_{xx}(1) \\ R_{xx}(1) & 1 \end{bmatrix}.$$

Обратная ей матрица

$$J^{-1} = \frac{b^2}{Ex_t^2} \cdot \frac{1}{n} \begin{bmatrix} 1 & -R_{xx}(1) \\ -R_{xx}(1) & 1 \end{bmatrix} \frac{1}{1-R_{xx}^2(1)},$$

где $R_{xx}(1) = Ex_t x_{t-1} / Ex_t^2$.

Среднее квадратичное значение ошибки характеризуется следом матрицы $R_{\hat{A}}$:

$$SpR_{\hat{A}} \geq \frac{1}{n} \cdot \frac{b^2}{Ex_t^2} \cdot \frac{2}{1-R_{xx}^2(1)}. \tag{22}$$

Из уравнения Юла-Уокера [6] имеем $R_{xx}(1) = a_1 / (1 - a_2)$, поэтому (22) можно записать и так:

$$SpR_{\hat{A}} \geq \frac{1}{n} \cdot \frac{b^2}{Ex_t^2} \cdot \frac{2(1-a_2)^2}{(1-a_2)^2 - a_1^2} \quad . \tag{23}$$

Сравнивая (22) и (19) , видим, что увеличение числа оцениваемых параметров увеличивает ошибку оценивания, однако с ростом n ошибку убывает так же, как $1/n$. Заметим, что ошибка (23) увеличивается даже если истинный параметр $a_2 = 0$. Это означает, то завышение порядка модели не желательно, так как ухудшает точность оценивания ее параметров.

Выводы

Из проведенного анализа видно, что на участках вокализированных звуков больше примерно на порядок, чем на участке невокализированных звуков даже при одинаковых значениях b . Это означает, что эффективная оценка параметров более точны, грубо говоря, на участках звучания гласных.

Список литературы

1. Дидковский В.С. – Акустическая экспертиза каналов речевой коммуникации / Дидковский В.С., Дидковская М.В., Продеус А.Н. – К.: Имэкс – ЛТД, 2008. – 420с.

2. Макаров Ю.К. – К оценке эффективности защиты акустической (речевой) информации / Макаров Ю.К., Хорев А.А. – (http//st.ess.ru/publications/tspi.htm).

3. Бузов Г.А. – Защита от утечки информации по техническим каналам / Бузов Г.А., Калинин С.В., Кондратьев А.В. – М.: Горячая линия – Телеком, 2005. – 414с.

4. Демьян Н.И. – Линейные модели речевого сигнала локально-постоянными параметрами / Демьян Н.И., Осмаловский В.А., Хорошко В.А. // Захист інформації, №1, 2010. – с.94-101.

5. Рабинер Л.Р. – Цифровая обработка речевых сигналов : Пер. с англ. / Рабинер Л.Р., Шафер Р.В. / Под ред. М.В. Назарова, Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 495с.

6. Карпов О.Н. – Компьютерные технологии распознавания речевых сигналов / Карпов О.Н., Габович А.Г., Марченко Б.Г., Хорошко В.А., Щербак Л.Н. – К.: ООО «ПолиграфКонсалтинг», 2005. – 138с.

В работе приведены модели человеческого голосового тракта (акустическая, электрическая и математическая) и на их основании проведен анализ участков вокализированных и невокализированных звуков.

Ключевые слова: модель речевых сигналов, оценка параметров, вокализированный и невокализированный звук.

В роботі запропоновано моделі голосового апарату людини (акустична, електрична та математична). На їх основі проведений аналіз ділянок вокалізованих та невокалізованих звуків.

Ключові слова: модель мовних сигналів, оцінка апарметрів, вокалізований та невокалізований звук.

Tha models of human vocal track (acoustic, electric and mathematical) are proposed in the article. On thear basis the analysis of sections of voiced and unvoiced sound has been made.

Key words: voice signal model, estimate of characteristics, voiced and unvoiced sound.

Поступила 24.02.2010