УДК 004.896

ОБНАРУЖЕНИЕ ЗВУКОВ РЕЧИ НА ФОНЕ ШУМОВ

Е.Г. ЖИЛЯКОВ
С.П. БЕЛОВ

Белгородский государственный национальный исследовательский университет

e-mail: Zhilyakov@bsu.edu.ru
e-mail: Belov@bsu.edu.ru

В статье теоретически обосновывается выбор решающей функции (РФ) для выявления на фоне шумов регистрирующих приборов факта присутствия в анализируемом отрезке речевого сигнала энергии, обусловленной наличием звуков речи. Показано, что предлагаемая РФ является функцией максимальной чувствительности в смысле приращения ее математического ожидания при наличии звуков речи.

Ключевые слова: частотные представления, решающая функция, звуки речи, минимизация объемов битовых представлений речевых данных, градиент.

Введение.

Известно, что речевые сигналы состоят в общем случае из отрезков отсчетов, которые сформировались либо при отсутствии звуков речи (шумы регистрирующих приборов в паузах речи) либо при наличии, как шумов, так и воздействий акустических колебаний, порождаемых звуками речи. Одна из задач проблемы минимизации объемов битовых представлений речевых данных заключается в сокращении данных в паузах речи на основе кодирования их длительностей, например, сохраняя номер начального отсчета и информацию об общем их количестве. При этом оказывается возможным почти на треть сократить объем хранимых и передаваемых данных. Эффективность реализации этого подхода в значительной степени зависит от правильного выбора решающей функции.

В связи с этим в статье обосновывается выбор решающей функции, позволяющей минимизировать вероятность ошибочного принятия решения.

Теоретические основы выбора решающей функции.

Уточним формулировку задачи обнаружения звуков речи применительно к анализу отрезков речевых сигналов.

Пусть зафиксирован отрезок (вектор) отсчетов речевого сигнала $\vec{x} = (x_1,..,x_N)^T$ длительности (размерности) N. Основная (начальная) гипотеза формулируется следующим образом:

H_0: отсчеты речевого сигнала принадлежат паузе в речи (порождена шумами).

Символически это означает, что

$$H_0 : \vec{x} = \vec{u} = (u_1,..,u_N)^T , \tag{1}$$

где символ u_i означает отсчет шума.

Формулировка альтернативной гипотезы имеет вид:

H_1: хотя бы часть отсчетов порождена, как шумами, так и под воздействием акустических колебаний, порождаемых звуками речи.

Полагая справедливым предположение об аддитивности взаимодействия шумов и реакции на звуковое воздействие, содержание альтернативы можно выразить следующим образом

$$H_1 : \vec{x} = \vec{z} + \vec{u} , \tag{2}$$

где $\vec{z} = (z_1,..,z_N)^T$ - обусловленный речевым сообщением вектор, некоторые из компонент которого могут быть равны нулю (например, когда речь отсутствует), причем предполагается что шумы от речи не зависят.

Необходимо разработать решающую процедуру, которая позволяет принять решение в пользу одной из сформулированных гипотез. При этом следует иметь в виду разную плату за ошибочные решения. В частности ошибочное решение в пользу гипотезы (2) приводит к неоправданному увеличению хранимых или передаваемых речевых данных. В свою очередь ошибочное решение в пользу гипотезы (1) приводит к искажению части речевого сообщения. Безусловно важнее исключить риски ошибочного принятия гипотезы (1).

Основу любой решающей процедуры составляет решающая функция (РФ):

$$F = F(\vec{x}), \tag{3}$$

которая определяет способ обработки вектора анализируемых данных, с тем чтобы при выполнения условия

$$F \notin D_\alpha \tag{4}$$

отвергнуть проверяемую гипотезу.

Здесь и в дальнейшем D_α - критическая область, выбираемая из условия

$$P\{F \notin D_\alpha / H_0\} = \alpha << 1, \tag{5}$$

где P - символ вероятности; α - вероятность ошибок первого рода.

Известно, что характеристическим свойством отрезков сигналов, порождаемых звуками речи, является сосредоточенность их энергий в малых частях частотной полосы, равной половине частоты дискретизации. Иными словами с высокой точностью выполняется условие

$$\| \vec{z} \|^2 = \sum_{k=1}^{N} z_k^2 \approx \sum_{r \in R_w} P_r(\vec{z}), \tag{6}$$

где $P_r(\vec{z})$ - части энергии

$$P_r(\vec{z}) = \int_{v \in V_r} |Z(v)|^2 \, dv / 2\pi, \tag{7}$$

попадающие в непересекающиеся частотные интервалы

$$V_r = [-V_{2r}, -V_{1r}) \bigcup [V_{1r}, V_{2r}), 0 \le \mathrm{V}_{1r} < V_{2r} \le \pi; V_n \bigcap V_m = \varnothing, \tag{8}$$

из некоторого множества R_w, причем их суммарная ширина удовлетворяет неравенству

$$\sum_{\partial \in R_w} (V_{2r} - V_{1r}) / \pi \le \gamma. \tag{9}$$

Здесь и в дальнейшем большими символами обозначаются трансформанты Фурье соответствующих векторов, обозначаемых малыми символами, то есть

$$Z(v) = \sum_{k=1}^{N} z_k \exp(-jv(k-1)). \tag{10}$$

Для правой части этого неравенства в зависимости от звука речи выполняется

$$0{,}25 \le \gamma \le 0{,}5, \tag{11}$$

причем правая часть здесь достигается редко (только для шипящих).

В свою очередь энергия отрезков шумов в паузах распределена более равномерно, так что предположительно выполняются равенства

$$b_r = M[P_r(\vec{u})] = M[\| \vec{u} \|^2](V_{2r} - V_{1r}) / \pi. \tag{12}$$

Здесь *М*- символ математического ожидания.

Поэтому, в основе построения РФ целесообразно использовать характеристику распределения энергий анализируемого отрезка сигнала по частотным интервалам

$$P_r(\vec{x}) = \int_{v \in V_r} |X(v)|^2 \, dv / 2\pi. \tag{13}$$

Имея в виду определения (3.10) и (3.8), отсюда нетрудно получить следующее представление в виде квадратичной формы

$$P_r(\vec{x}) = \vec{x}^T A_r \vec{x}, \tag{14}$$

где $A_r = \{a_{ik}^r\}, i,k = 1,..,N$ - субполосная матрица с элементами

$$a_{ik}^r = \int_{v \in V_r} \exp(-jv(i-k))dv / 2\pi = 2\sin(\Delta V_r (i-k)/2) * \cos(\omega_r (i-k)) / \pi(i-k); \quad (15)$$

$$\Delta V_r = V_{2r} - V_{1r}; \quad (16)$$

$$\omega_r = (V_{2r} + V_{1r})/2; \quad (17)$$

Здесь в дальнейшем символ T в качестве верхнего индекса означает транспонирование соответствующего вектора.

Пусть для определенности отсчеты шумов в паузах речи являются стационарными, имеют равное нулю математическое ожидание

$$M[u_k] = 0, k = 1,..,N, \quad (18)$$

и автокорреляционную функцию (АКФ)

$$G_u(i-k) = M[u_k u_i],\ i,k = 1,..,N. \quad (19)$$

Тогда из определения (14) нетрудно получить выражение для условных математических ожиданий соответствующих частей энергий отрезков шума

$$M[P_r(\vec{x})/H_0] = b_r = M[P_r(\vec{u})] = \sum_{i,k=1}^{N} a_{ik}^r G_u(i-k). \quad (20)$$

Если шумы не коррелированны

$$\begin{aligned} G_u(i-k) &= \sigma_u^2, i = k; \\ G_u(i-k) &= 0,\ \mathrm{i} \neq \mathrm{k} \end{aligned} \quad (21)$$

то с учетом соотношений (15) получаем

$$M[P_r(\vec{x})/H_0) = b_r = M[P_r(\vec{u})] = \sigma_u^2 N \Delta V_r / \pi. \quad (22)$$

В свою очередь нетрудно получить представление для условного математического ожидания квадрата доли энергии

$$M[P_r^2(\vec{x})/H_0) = M[P_{\grave{e}}^2(\vec{u})] = \sum_{i,k=1}^{N} \sum_{n,m=1}^{N} a_{ik}^r a_{nm}^r M[u_i u_k u_n u_m]. \quad (23)$$

Предположим также, что вектор шумов имеет гауссовское распределение вероятностей и выполняются условия (18) и (21). Тогда [1] справедливы соотношения

$$M[u_i u_k u_n u_m] = 0,\ \ (i \neq k) \cap (i \neq n) \cap (i \neq m) \cap (k \neq n) \cap (k \neq m) \cap (n \neq m)$$

$$M[u_i u_k u_n u_m] = 0, \quad (24)$$

если хотя бы один из индексов не совпадает ни с одним из других;

$$M[u_i u_k u_n u_m] = \sigma_r^2\ \ \sigma_{\mathrm{r}}^2, \quad (25)$$

если имеются попарные совпадения индексов;

$$M[u_i u_k u_n u_m] = 3\sigma_r^4, i = k = n = m. \quad (26)$$

Таким образом, соотношение (23) дает

$$M[P_r^2(\vec{x})/H_0) = M[P_{\grave{e}}^2(\vec{u})] = (\sigma_u^2 N \Delta V_r / \pi)^2 + \sigma_u^4 N (\Delta V_r / \pi)^2. \quad (27)$$

Отсюда с учетом соотношения (22) нетрудно получить представление для условной дисперсии рассматриваемой части энергии

$$d_r^2 = M[P_r^2(\vec{x})/H_0)] - M^2[P_r(\vec{x})/H_0] = 2\sigma_u^4 N(\Delta V_r/\pi)^2. \tag{28}$$

Представления (22) и (28) дают соотношение

$$\varphi_r = d_r / b_r = (2/N)^{1/2}, \tag{29}$$

которое в оговоренных выше условиях определяет закономерность изменений отношения среднеквадратического отклонения части энергии отрезка шума к его математическому ожиданию. Легко видеть, что с ростом длительности анализируемого отрезка сигнала в паузе это отношение стремиться к нулю.

Исследование свойств частей энергий шума можно осуществить и с иных позиций, имея в виду, что симметричная субполосная матрица обладает полным набором ортонормированных собственных векторов [2], удовлетворяющих условиям

$$\lambda_k^r \vec{q}_k^r = A_r \vec{q}_k^r,\ k = 1,..,N; \tag{30}$$

$$\lambda_1^r > \lambda_2^r > \ldots > \lambda_N^r; \tag{31}$$

$$(\vec{q}_k^r, \vec{q}_i^r) = \sum_{n=1}^{N} q_{nk}^r q_{ni}^r = \delta_{ik},\ i,k = 1,..,N; \tag{32}$$

δ_{ik} - символ Кронекера

$$\begin{matrix} \delta_{ik} = 0\ \forall\, i \neq k, \\ \delta_{øё} = 1\ \forall i = k \end{matrix}.$$

Отметим, что, как показывают вычисления, собственные числа субполосных матриц обладают свойством

$$\lambda_{J+n}^r \approx 0,\ n = 1,..,N-J,\ J = 2[\Delta V_r / 2\pi] + 4. \tag{33}$$

Тогда [2] субполосную матрицу можно представить в виде

$$A_r = \sum_{k=1}^{N} \lambda_k^r \vec{q}_k^r \vec{q}_k^{rT}, \tag{34}$$

что позволяет вычислить часть энергии в заданном частотном интервале с использованием соотношения

$$P_r(\vec{x}) = \sum_{k=1}^{N} \lambda_k^r \alpha_{kr}^2, \tag{35}$$

где α_{kr} - скалярные произведения анализируемого и соответствующего собственного векторов

$$\alpha_{kr} = (\vec{x}, \vec{q}_k^r) = \sum_{i=1}^{N} x_i q_{ik}^r, \tag{36}$$

В виду линейности формы (36) при выполнении гипотезы H_0 и оговоренных выше свойствах шумов в паузах речи эти случайные величины также будут иметь гауссовское распределение вероятностей. При этом с учетом ортонормальности собственных векторов субполосных матриц нетрудно получить соотношения для первых двух условных моментов

$$M[\alpha_{kr} / H_0] = 0, \tag{37}$$

$$M[\alpha_{kr}\alpha_{ir} / H_0] = 0,\ i \neq k, \tag{38}$$

$$M[\alpha_{kr}^2 / H_0] = \sigma_u^2. \tag{39}$$

В соответствии с этим соотношение (35) дает

$$M[P_r(\vec{x})/H_0] = \sigma_u^2 \sum_{k=1}^{N} \lambda_k^r. \tag{40}$$

Имея в виду связи между следом матрицы и суммой её собственных чисел [2]

$$\sum_{k=1}^{N} \lambda_k^r = \sum_{k=1}^{N} a'_{kk}, \tag{41}$$

и определение (3.15), отсюда снова нетрудно получить соотношение (3.22). Кроме того с учетом гауссовости случайных величин вида (3.36) и равенств (3.37)-(3.39) нетрудно получить соотношение для условной дисперсии

$$M[(P_r(\vec{x}) - b_r)^2 / H_0] = 2\sigma_u^4 \sum_{k=1}^{N} (\lambda_k^r)^2. \tag{42}$$

Сопоставляя соотношения (40) и (44) с полученными ранее из других соображений представлением (22) и (28)) первых двух условных моментов части энергии с учетом равенства (41) нетрудно установить следующие равенства

$$\sum_{k=1}^{N} (\lambda_k^r)^2 = N(\Delta V_r / \pi)^2 = (\Delta V / \pi) \sum_{k=1}^{N} \lambda_k^r. \tag{43}$$

Вместе с тем, если учесть, что только часть собственных чисел субполосной матрицы значимо отлична от нуля, то соотношения (35) ,(40) и (42) нетрудно преобразовать к виду

$$P_r(\vec{x}) = \sum_{k=1}^{J} \lambda_k^r \alpha_{kr}^2, \tag{44}$$

$$b_r = M[P_r(\vec{x}) / H_0] = \sigma_u^2 \sum_{k=1}^{J} \lambda_k^r. \tag{45}$$

$$M[(P_r(\vec{x}) - b_r)^2 / H_0] = 2\sigma_u^4 \sum_{k=1}^{J} (\lambda_k^r)^2. \tag{46}$$

В виду гауссовости при выполнении гипотезы H_0 случайных величин (36), и, следовательно, в силу (38) их независимости можно показать [1], что характеристическая функция случайной величины (44) имеет вид

$$Q(t / H_0) = \prod_{k=1}^{J} (1 - j\lambda_k^r \sigma_u^2 t). \tag{47}$$

Дифференцируя это выражение, нетрудно по известным правилам [1] получить соотношения для любых условных начальных моментов. В частности имеет место представление для условной функции плотности вероятностей вычисляемой части энергии

$$w_P(y / H_0) = \int_{-\infty}^{\infty} \prod_{k=1}^{J} (1 - j\sigma_u^2 \lambda_k^r t) \exp(-jyt) dt, \tag{48}$$

которое в принципе позволяет вычислить границы критической области, исходя из заданной вероятности ошибок первого рода при проверке справедливости гипотезы H_0.

Положим

$$S(\omega_r, \vec{x}) = P_r(\vec{x}) / M[P_r(\vec{x}) / H_0] = P_r(\vec{x}) / b_r. \tag{49}$$

Используя предыдущие результаты, легко показать справедливость соотношений

$$c_r^0 = M[S(\omega_r, \vec{x}) / H_0] = 1, \tag{50}$$

$$\sigma_S^2 = M[(S_r(\vec{x}) - 1)^2 / H_0] = 2 / N. \tag{51}$$

Отметим, что соотношение (51) характеризует закон изменения отклонений (49) относительно правой части (50) в зависимости от длительности анализируемого отрезка шумов в паузах речи.

В общем случае (2) имеет место

$$S(\omega_r, \vec{x}) = (P_r(\vec{u}) + \vec{z}^T A_r \vec{u} + P_r(\vec{z})) / b_r . \tag{52}$$

Отсюда с учетом предыдущих условий и результатов получаем

$$c_r^1 = M[S(\omega_r, \vec{x}) / H_1] = 1 + M[P_r(\vec{z})] / b_r . \tag{53}$$

Так как величина интеграла (7) при ненулевом векторе сигнала является положительной, то отсюда следует неравенство

$$c_r^1 > c_r^0 = 1 . \tag{54}$$

Иными словами, математическое ожидание случайной величины (52) увеличивается по сравнению с выполнением исходной гипотезы.

Пусть теперь полезный сигнал в (2) является узкополосным и может быть представлен в виде

$$z_k = y_k \cos(\theta k + \varphi), k = 1,.., N , \tag{55}$$

где φ - случайная равномерно распределенная в интервале $(-\pi, \pi)$ фаза; θ - круговая частота из интервала $(0, \pi)$; y_k - так называемая огибающая.

Полагаем, что величина $N(\Delta V_r / \pi)$ выбрана такой, что выполняется условие

$$P_0(\vec{y}) = 2 \int_{0_r}^{\Delta V_r / 2} | Y(v) |^2 \, dv / 2\pi = m \sum_{k=1}^{N} y_k^2 , \ m \approx 1 , \tag{56}$$

причем имеет место представление

$$P_0(\vec{y}) = \vec{y}^T A_0 \vec{y} = \sum_{i,k=1}^{N} a_{ik}^0 y_i y_k , \tag{57}$$

$$a_{ik}^0 = \sin(\Delta V_r (i-k) / 2) / \pi (i-k) \tag{58}$$

Очевидно, что ввиду равенства[1]

$$M[z_i z_k] = y_i y_k \cos(\theta (i-k)) \tag{59}$$

будет выполняться

$$M[P_r(\vec{z})] = 2 \sum_{i,k=1}^{N} a_{ik}^0 y_i y_k \cos(\omega_r (i-k)) \cos(\theta (i-k)) . \tag{60}$$

Отсюда получаем выражение для первых двух производных

$$dM[P_r(\vec{z})] / d\omega_r = -2 \sum_{i,k=1}^{N} (i-k) a_{ik}^0 y_i y_k \sin(\omega_r (i-k)) \cos(\theta (i-k)) , \tag{61}$$

$$d^2 M[P_r(\vec{z})] / d\omega_r^2 = -2 \sum_{i,k=1}^{N} (i-k)^2 a_{ik}^0 y_i y_k \cos(\omega_r (i-k)) \cos(\theta (i-k)) , \tag{62}$$

из которых следует, что вторая производная всюду неположительная, так что в максимум правой части (60) будет достигаться там, где равна нулю правая часть (61), то есть выполняется

$$2 \sum_{i=1}^{N} i y_i \sum_{k=1}^{N} a_{ik}^0 y_i \sin(\omega_r (i-k)) \cos(\theta (i-k)) = 2 \sum_{k=1}^{N} k y_k \sum_{i=1}^{N} a_{ik}^0 y_i \sin(\omega_r (i-k)) \cos(\theta (i-k)).$$

Легко понять, что в виду симметрии субполосной матрицы, четности функции косинуса и нечетности функции синуса последние условие равносильно требованию

$$2 \sum_{i=1}^{N} i y_i \sum_{k=1}^{N} a_{ik}^0 y_i \sin(\omega_r (i-k)) \cos(\theta (i-k)) = 0 . \tag{63}$$

Его легко на основе тригонометрических тождеств преобразовать к виду

$$\sum_{i=1}^{N} iy_i \sum_{k=1}^{N} a_{ik}^0 y_i \sin((\omega_r - \theta)(i-k)) = -\sum_{i=1}^{N} iy_i \sum_{k=1}^{N} a_{ik}^0 y_i \sin((\omega_r + \theta)(i-k)) \,. \tag{64}$$

В частном случае равенства частот

$$\omega_r = \theta \tag{65}$$

получаем требование в виде

$$\sum_{i=1}^{N} iy_i \sum_{k=1}^{N} a_{ik}^0 y_k \sin((2\theta k - 2\theta i)) = 0 \,. \tag{66}$$

Так спектр последовательности $y_k \sin((2\theta k - 2\theta i))$ по сравнению со спектром огибающей в (55) будет сдвинут на 2θ в область высоких частот, то в виду условия (56) равенство (63) будет выполняться с той же точностью. Иными словами, выполнение равенства (62) соответствует достижению максимального значения правой части (60)

$$\max M[P_r(\vec{z})] = \sum_{i,k=1}^{N} a_{ik}^0 y_i y_k = P_0(\vec{y}), \tag{67}$$

подстановка которого в (53) в свою очередь дает значение максимума математического ожидания отношения вида (52)

$$\max c_r^1 = 1 + P_0(\vec{y}) / b_r \,, \tag{68}$$

Таким образом, математическое ожидание отношения

$$S(\omega_r, \vec{x}) = P_r(\vec{x}) / b_r \,, \tag{69}$$

увеличивается при наличии квазигармонической компоненты, обусловленной речевыми сообщениями (речевая компонента). При этом совпадение значения середины частотного интервала, в котором оценивается часть энергии, с частотой квазигармонической речевой компоненты дает максимум.

Пусть теперь речевая компонента представляет собой аддитивную смесь нескольких квазигармонических составляющих

$$z_k = \sum w_{ki} = \sum_{i=1}^{M} y_{ki} \cos(\theta_i k + \varphi_i), k = 1,..,N \,, \tag{70}$$

смысл параметров которой очевиден. Тогда будет иметь место

$$c_r^1 = M[S(\omega_r, \vec{x}) / H_1] = 1 + \sum_{i=1}^{M} M[P_r(\vec{w}_i)] / b_r > 1 \,. \tag{71}$$

При выполнении какого – нибудь из равенств

$$\omega_r = \theta_j, j \in \{1,..,M\} \tag{72}$$

будет достигать максимального значения соответствующее математическое ожидание в сумме правой части (71). Ясно, что при выполнении условия

$$\min |\theta_j - \theta_k| > \Delta V_r, j,k \in \{1,..,M\} \tag{73}$$

можно определить такую квазигармоническую $\vec{w}_m$ компоненту, для которой имеет место

$$P_0(\vec{y}_m) > P_0(\vec{y}_k), \forall k \neq m \,. \tag{74}$$

Иными словами, правая часть (69) может использоваться в качестве решающей функции, причем решение о наличии речевой компоненты должно приниматься при выполнении условия

$$\max S(\omega_r, \vec{x}) > h(\omega_r), \Delta V_r \leq \omega_r \leq \pi - \Delta V_r \,. \tag{75}$$

Здесь $h(\omega_r)$ - в общем случае зависящая от центральной частоты интервала оценивания частей энергии функция, обеспечивающая допустимые вероятности ошибок первого и второго родов. Её значения целесообразно вместе с функцией вида (22) (зависимость математического ожидания шумов от центральной частоты) оценивать непосредственно по реализациям шумов, для чего необходимо иметь достаточно обширную обучающую выборку. Отметим еще одно качество предлагаемой решающей функции. Определение (57) с учетом возможности разложения

$$A_0 = \sum_{k=1}^{N} \lambda_k^0 \vec{q}_k^0 \vec{q}_k^{0T}$$

нетрудно преобразовать к виду

$$P_0(\vec{y}) = D(\vec{\beta}) = \sum_{k=1}^{N} \beta_k^2,\ \beta_k = (\lambda_k^0)^{1/2} \sum_{m=1}^{N} y_m q_{mk}^0 . \qquad (76)$$

Отсюда после частного дифференцирования по β_n, $n = 1,..,N$ легко получить соотношение для градиента части энергии, попадающей в интервал частотного анализа

$$gradP = 2(\beta_1,..,\beta_N)^T , \qquad (77)$$

так что имеет место равенство

$$P_0(\vec{y}) = D(\vec{\beta}) = (gradP, \vec{\beta})/2 , \qquad (78)$$

Иными словами приращение (по сравнению с отсутствием сигнала) функции происходит в направлении градиента, то есть в линейном приближении максимально быстро. Таким образом, предлагаемая решающая функция в определенном смысле является функцией максимальной чувствительности.

Работа выполнена при поддержке гранта РФФИ 12-07-00514-а.

Литература

1. Левин, Б.Р.Теоретические основы статистической радиотехники [Текст] – М.: Советское радио, т.1, 1969. – 752 с.
2. Гантмахер, Ф.Р. Теория матриц [Текст] / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560с.

DETECTION OF SPEECH SOUNDS IN THE BACKGROUND NOISE

E.G. ZHILYAKOV
S.P.BELOV

Belgorod National Research University

e-mail: Zhilyakov@bsu.edu.ru
e-mail: Belov@bsu.edu.ru

The paper theoretically justified choice of decision function (RF) to identify the background noise recording apparatus of the fact of presence in the analyzed segment of the speech signal energy due to the presence of speech sounds. It is shown that the proposed RF is a function of maximum sensitivity in the sense of expectation of the increment in the presence of speech sounds

Key words: frequency representation, decision function, speech sounds, minimizing the volume of voice data bit representations, the gradient.