

О ВЫБОРЕ МЕТОДА АКУСТИЧЕСКОЙ ЭКСПЕРТИЗЫ И ВИДА ТЕСТОВОГО СИГНАЛА ПРИ ОЦЕНИВАНИИ РАЗБОРЧИВОСТИ РЕЧИ

А. Н. ПРОДЕУС

*Национальный технический университет Украины «КПИ», Киев
e-mail: aprodeus@gmail.com*

Предложены решающие правила, необходимые для автоматизации акустической экспертизы. Экспериментально оценены погрешности акустической экспертизы, обусловленные использованием естественной речи в качестве тестового сигнала.

ВВЕДЕНИЕ

В последние десятилетия широкое распространение получил модуляционный метод акустической экспертизы, универсальность которого обусловлена способностью учитывать влияние на разборчивость речи как шума, так и реверберации [1]. Между тем, в условиях преобладания шумовой помехи предпочтительным, как по скорости, так и по точности измерений, является формантный метод [2]. Отсюда следует, что выбор рационального метода акустической экспертизы целесообразно производить с учетом оценки соотношения влияния шума и реверберации. Поэтому одна из целей данной работы состоит в разработке вариантов решающего правила о виде помехи.

Другим важным вопросом акустической экспертизы является выбор тестового сигнала. При экспертизе помещений принято использовать тестовые сигналы в виде стационарного или нестационарного шумов, свип-сигналы или сигналы в виде MLS-последовательностей [3]. Однако специфика цифровых линий связи приводит к необходимости использовать естественную речь в качестве тестовых сигналов. До настоящего времени вопрос точности таких измерений исследован недостаточно, поэтому другой целью данной работы является оценка погрешностей акустической экспертизы, производимой с использованием речевых тестовых сигналов.

1. ВЫБОР МЕТОДА АКУСТИЧЕСКОЙ ЭКСПЕРТИЗЫ

В работе [2] показано, что в случае преобладания шумовой помехи наиболее точным и быстрым является формантный метод оценивания разборчивости речи. Однако при сопоставимом влиянии шумовой и реверберационной помех вместо формантного метода целесообразно применять ФМ метод, позволяющий корректно учитывать влияние реверберационной помехи на разборчивость речи. В случае преобладания реверберационной помехи весьма экономичной в вычислительном плане является разновидность ФМ метода, при которой коэффициент модуляции вычисляют по формуле Шредера:

$$m_{ik} \approx 2 \int_0^{\infty} h_{ek}^2(t) e^{-j2\pi F_i t} dt \Big/ \int_0^{\infty} h_{ek}^2(t) dt, \quad h_{ek}(v) = \int_{-\infty}^{\infty} h(v) h_{k0}(z-v) dv,$$

где $h_{k0}(v)$ – ИХ k -го полосового фильтра; $h(t)$ – ИХ помещения; F_i – частота модуляции тестового сигнала ($F_i = 0,63, \dots, 12,5$ Гц). При этом, если требуется обеспечить

скритність измерений, в качестве тестовых сигналов целесообразно использовать MLS-последовательности [3]. Схема процедуры выбора метода акустической экспертизы, базирующаяся на использовании таких критериев как точность, скорость и скрытность, с учетом возможного преобладания того или иного вида помехи, приведена на рис. 1.

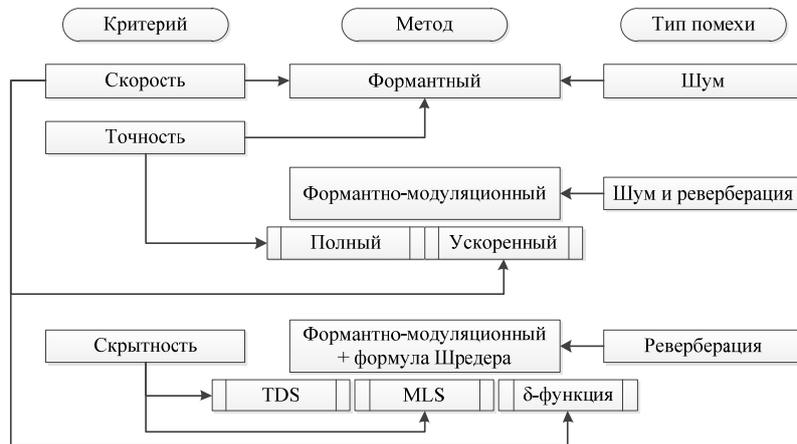


Рис. 1

Однако использование схемы рис. 1 неизбежно порождает вопрос: как установить характер помехи? Ответить на данный вопрос сравнительно легко, если использовать понятие «эффективного отношения сигнал-шум» [4].

Действительно, поскольку отношение сигнал-шум SNR_N для шумовой помехи связано с коэффициентом модуляции m_N соотношением

$$SNR_N = 10 \lg \frac{m_N}{1 - m_N}, \quad (1)$$

«эффективное» отношение сигнал-шум SNR_R для реверберации определяют через коэффициент модуляции m_R в соответствии с аналогичным соотношением [1]:

$$SNR_R = 10 \lg \frac{m_R}{1 - m_R}, \quad m_R = \left[1 + \left(\frac{2\pi F_i T_{60}}{13,8} \right)^2 \right]^{-0,5}, \quad (2)$$

где $F_i = 0,63 \dots 12,5$ Гц – частота модуляции тестового сигнала; T_{60} – время реверберации.

Отсюда следует удобное для инженерных приложений правило принятия решения о типе помехи (рис. 2):

- преобладает шум, если $SNR_R - SNR_N \geq 10 \dots 15$ дБ;
- преобладает реверберация, если $SNR_R - SNR_N \leq -10 \dots 15$ дБ;
- влияние сопоставимо, если $-15 \dots -10$ дБ $\leq SNR_R - SNR_N \leq 10 \dots 15$ дБ.

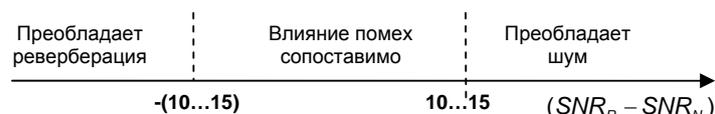


Рис. 2

Используя соотношения (1)-(2), нетрудно получить уравнение, связывающее между собой m_R и m_N для соответствующих значений разницы $SNR_R - SNR_N$:

$$m_N = [10^{0,1(SNR_R - SNR_N)} (m_R^{-1} - 1) + 1]^{-1}. \quad (3)$$

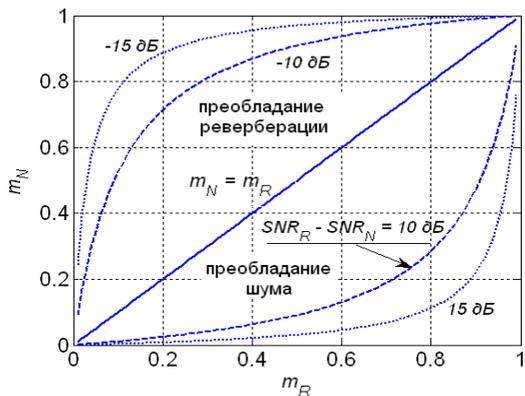


Рис.3

Полагая $SNR_R - SNR_N = 10 \dots 15$ дБ либо $SNR_R - SNR_N = -(10 \dots 15)$ дБ, получаем верхние граничные линии (пунктирную и точечную), между которыми находится область значений m_N и m_R , соответствующая сопоставимому влиянию шума и реверберации (рис. 3).

Таким образом, рис. 3 графически представляет альтернативную версию правила принятия решение о характере помех:

- преобладает шум, если $m_N \leq [10^{0,1(10 \dots 15)} (m_R^{-1} - 1) + 1]^{-1}$;
- преобладает реверберация, если $m_N \geq [10^{-0,1(10 \dots 15)} (m_R^{-1} - 1) + 1]^{-1}$;
- влияние сопоставимо, если $[10^{0,1(10 \dots 15)} (m_R^{-1} - 1) + 1]^{-1} < m_N < [10^{-0,1(10 \dots 15)} (m_R^{-1} - 1) + 1]^{-1}$.

Если при измерениях оценивается время реверберации T_{60} и отношение сигнал-шум SNR_N для шумовой помехи, тогда целесообразно использовать третью версию правила принятия решения:

- преобладает шум, если $SNR_N \leq 10 \lg m_R / (1 - m_R) - 10 \dots 15$ дБ;
- преобладает реверберация, если $SNR_N \geq 10 \lg m_R / (1 - m_R) + 10 \dots 15$ дБ;
- влияние сопоставимо, если $10 \lg m_R / (1 - m_R) + 10 \dots 15 \leq SNR_N \leq 10 \lg m_R / (1 - m_R) - 10 \dots 15$ дБ.

Графически эта версия представлена на рис. 4 и 5 для частот модуляции $F_i = 1$ Гц и $F_i = 10$ Гц, соответственно.

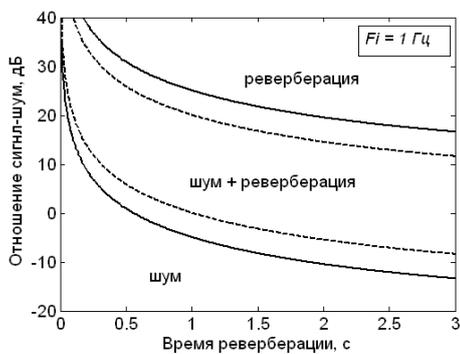


Рис. 4

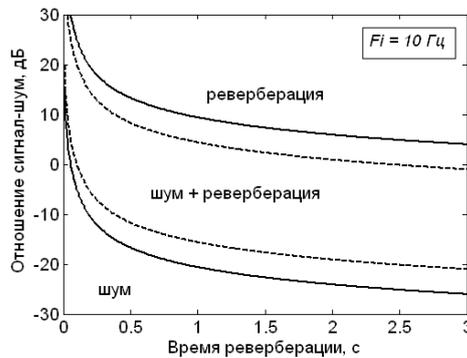


Рис. 5

2. ИСПОЛЬЗОВАНИЕ РЕЧИ В КАЧЕСТВЕ ТЕСТОВОГО СИГНАЛА

Тестовые сигналы в виде стационарного или нестационарного шума, обычно используемые при акустической экспертизе помещений, малопригодны для сквозного оценивания качества типичных для систем связи алгоритмов обработки сигналов, таких как алгоритмы шумоподавления и кодирования. Поясняется это тем, что разборчивость речи важна преимущественно для военных линий связи, использующих вокодеры. Между тем, для коммерческих линий связи, а также для линий связи с автоматическим распознаванием речи, не менее важную роль играет такой показатель как качество речи [4]. Поэтому в последнее время значительно усилился интерес исследователей к использованию естественной речи в качестве тестовых сигналов, позволяющих оценить как качество, так и разборчивость передаваемой речи [5, 6]. При этом особо следует выделить три способа оценивания коэффициента передачи модуляции m_{ik} .

Первый способ, «классический», состоит в оценивании m_{ik} через спектры мощности $S_{XX}(f)$ и $S_{YY}(f)$ огибающих $X(t)$ и $Y(t)$ тестового сигнала $x(t)$ и принимаемого сигнала $y(t)$:

$$m_{ik} = \alpha \sqrt{\frac{S_{kYY}(F_i)}{S_{kXX}(F_i)}}, \quad \alpha = \frac{M\{X(t)\}}{M\{Y(t)\}}, \quad (4)$$

где $M\{f(t)\}$ - оператор вычисления среднего значения функции $f(t)$.

Другой способ, основанный на использовании взаимного спектра мощности $S_{XY}(f)$ огибающих $X(t)$ и $Y(t)$, предложен в работе К. Пэйтона [5]:

$$m_{ik} = \alpha \frac{|S_{kXY}(F_i)|}{|S_{kXX}(F_i)|}, \quad (5)$$

где $||$ - символ модуля комплексного числа. Р. Друллман в работе [6] исследовал третий способ оценивания, базирующийся на соотношении:

$$m_{ik} = \alpha \frac{\operatorname{Re}\{S_{kXY}(F_i)\}}{S_{kXX}(F_i)}, \quad (6)$$

где $\operatorname{Re}\{A\}$ - оператор выделения действительной части комплексного числа A .

К сожалению, результаты упомянутых работ не сопоставлены между собой для ситуаций шумов различной окрашенности, а также не содержат данных о методической погрешности каждого из способов. Ниже представлены результаты экспериментальных исследований, направленных на восполнение указанного пробела.

При проведении экспериментальных исследований квадрат огибающей $X(t)$ узкополосного сигнала $x(t) = X(t) \cos(2\pi f_0 t + \varphi(t))$ ($\varphi(t)$ - фаза сигнала $x(t)$; f_0 - частота гармонической несущей) вычислялся двумя способами. Первый способ базировался на использовании системы «квадратор-интегратор»:

$$X^2(t) \approx \frac{2}{T_0} \int_{t-T_0}^t [X(t) \cos(2\pi f_0 t + \varphi(t))]^2 dt,$$

где время интегрирования $T_0 = 0,01$ с выбрано из условия $1/f_0 \ll T_0 \ll 1/F_i$ ($f_0 = 125, \dots, 8000$ Гц). Второй способ основывался на использовании преобразования Гильберта $\hat{x}(t)$ узкополосного сигнала $x(t)$: $X^2(t) = x^2(t) + \hat{x}^2(t)$. Поскольку функция

$X^2(t)$ медленная, по сравнению с функцией $x(t)$, можно существенно, почти на два порядка (с 22 050 Гц до 250 Гц), снизить частоту ее дискретизации, что позволяет почти на два порядка снизить объем последующих вычислений.

На рис. 6 представлены результаты вычислений словесной разборчивости речи (четкая речь в умеренном темпе, диктор - мужчина) в условиях действия шумовой помехи для двух видов тестового сигнала: шумового (рис. 6,а) и речевых сигналов (рис. 6,б – способ 1; рис. 6,в – способ 2; рис. 6,г – способ 3).

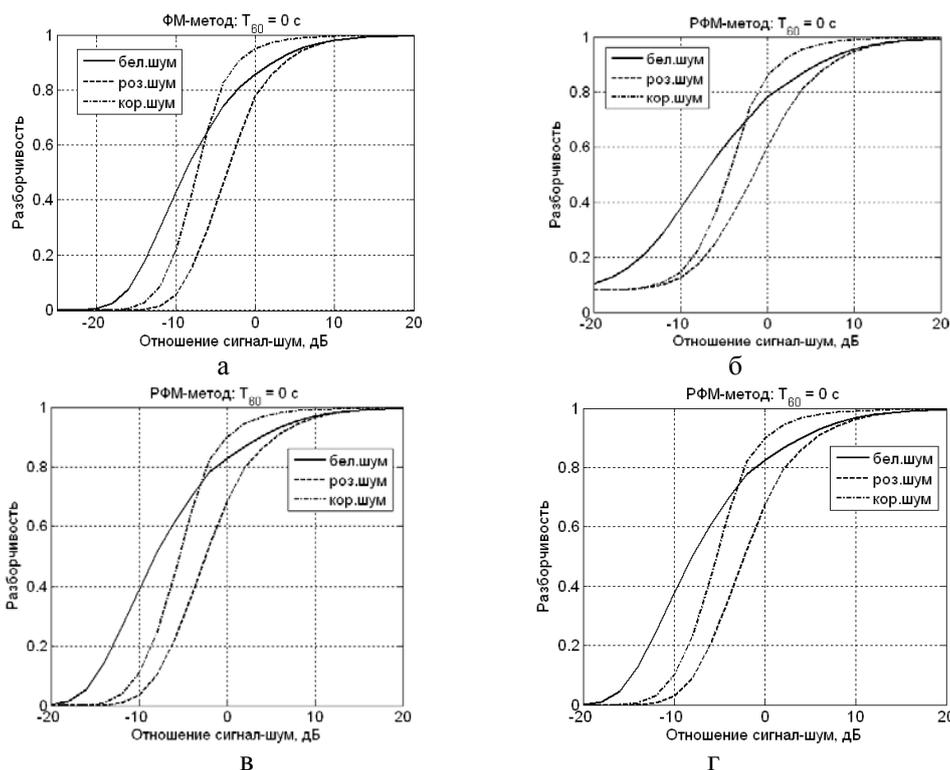


Рис. 6

Сопоставление приведенных на рис. 6 графиков свидетельствует, что абсолютная погрешность оценивания словесной разборчивости речи максимальна для способа 1 и превышает 0,2, тогда как для способов 2 и 3 эта погрешность не превышает 0,1.

На рис. 7 представлены аналогичные результаты вычислений словесной разборчивости речи в условиях совместного действия шума и реверберации (время реверберации $T_{60} = 2,9 с$).

Сопоставление приведенных на рис. 7 графиков свидетельствует, что в условиях интенсивной реверберации наилучшие результаты обеспечивает способ 1 (абсолютная погрешность измерений не превышает 0,1), тогда как способы 2 и 3 приводят к неприемлемо большим погрешностям (0,2...0,6). В заключение отметим еще один существенный недостаток речи как тестового сигнала: результаты акустической экспертизы зависят не только от времени анализа, темпа и четкости речи, но и от диктора.

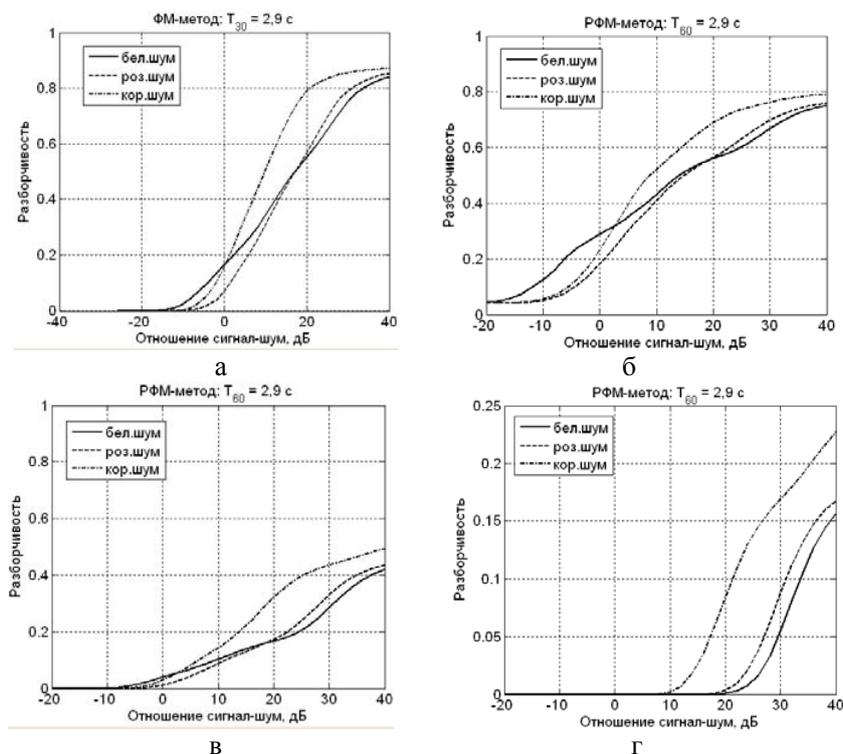


Рис. 7

ВЫВОДЫ

Предложены несколько вариантов решающего правила о виде помехи, необходимого для автоматизации акустической экспертизы, а также экспериментально оценены погрешности измерения словесной разборчивости речи при использовании естественной речи в качестве тестового сигнала. Полученные результаты позволяют управлять точностью и уровнем автоматизации акустической экспертизы помещений и линий связи.

ЛИТЕРАТУРА

1. *Springer Handbook of Acoustics*. – New York, 2007. – 1182 p.
2. Продеус А.Н. Формантно-модуляционный метод оценки разборчивости речи: точность и скорость измерений. - Сб.трудов Акустического симпозиума "Консонанс-2011", К., 2011, с.224-229.
3. Stan G.-B., Embrechts J.-J., Archambeau D. Comparison of Different Impulse Response Measurement Techniques. JAES Volume 50 Issue 4 pp. 249-262; April 2002.
4. Recommendation ITU-T P.863. Methods for objective and subjective assessment of speech quality. Switzerland Geneva, 2011. 76 p.
5. Payton, K.L., Braida, L.D., Chen, S. Rosengard, P. and Goldsworthy, R. Computing the STI using speech as a probe stimulus. In Past, Present and Future of the Speech Transmission Index., (ed. van Wijngaarden), TNO Human Factors: Soesterberg, the Netherlands, (2002) 125-137.
6. Drullman R., Festen J., Plomp R. Effect of reducing slow temporal modulations on speech receptions. – JASA, № 95, 1994, P. 2670-2680.