

УДК 534.782.001:621.39

А.Н. Продеус

Національний технічний університет України «КПІ», Київ

РЕЧЕВОЙ ТЕСТОВЫЙ СИГНАЛ В ФОРМАНТНО-МОДУЛЯЦИОННОМ МЕТОДЕ ОЦЕНИВАНИЯ РАЗБОРЧИВОСТИ РЕЧИ

Рассмотрена принципиальная возможность применения тестовых сигналов в виде естественной речи в формантно-модуляционном методе оценивания разборчивости речи. Показано, что способ оценивания коэффициента передачи модуляции, основанный на спектральном анализе интенсивностей огибающих тестового и выходного сигналов, обеспечивает получение правдоподобных результатов в условиях действия шумовой помехи, однако в условиях интенсивной реверберации данный способ практически не работоспособен.

Ключевые слова: оценивание разборчивости речи, формантно-модуляционный метод, коэффициент передачи модуляции, шумовая помеха, реверберация.

Введение

Варианты классического формантного метода оценивания разборчивости речи, описанные в работах Н.Б. Покровского, М.А. Сапожкова и Ю.С. Быкова, отличаются, главным образом, трактовкой понятия «формантный спектр» и наборами коэффициентов восприятия. Причины различия этих коэффициентов восприятия проанализированы в работе [1]. Экспериментально измеренные, на этой теоретической основе, уточненные коэффициенты восприятия для русской и украинской речи представлены в работе [2].

К сожалению, формантный метод, изначально разработанный для оценки влияния шумовой помехи, не позволяет корректно учесть влияние реверберации на разборчивость речи. Поэтому в работе [3] был предложен формантно-модуляционный (ФМ) метод, свободный от этого недостатка. Экспериментальная проверка работоспособности, а также оценка потенциальных возможностей в смысле точности и скорости вычислений ФМ метода осуществлены в работах [4, 5]. Поскольку для вычислений, реализующих ФМ метод, требуется достаточно много времени, была разработана «быстрая» версия ФМ метода [6].

На практике при акустической экспертизе каналов речевой коммуникации обычно используют тестовые сигналы в виде стационарного (формантный метод) или нестационарного (ФМ метод) шума, спектр которого подобен долговременному спектру речевого сигнала. Между тем, такие тестовые сигналы малоприспособлены для оценивания качества алгоритмов обработки сигналов (шумоподавление, кодирование и синтез речи, и автоматическое распознавание речи) в системах связи. Поэтому в последнее время значительно усилился интерес исследователей к использованию естественной речи в качестве тестовых сигналов [7 – 9].

В данной работе рассмотрена принципиальная возможность расширения возможностей ФМ метода за счет использования речевых тестовых сигналов.

1. Основные аналитические соотношения ФМ метода

При акустической экспертизе помещений, характеризующихся импульсной характеристикой (ИХ) $h(v)$, в определенной точке (местоположение диктора) излучают тестовый сигнал $x(t)$, а в иной точке (местоположение аудитора) принимают сигнал

$$y(t) = \int_{-\infty}^{\infty} h(v)x(t-v)dv + n(t),$$

где $n(t)$ – шумовая помеха.

При использовании ФМ метода [4, 5], как и в формантном методе [1], вычисляют вероятность правильного понимания формант:

$$A = \sum_{k=1}^K p_k \cdot P_k(E_k), \quad (1)$$

где p_k – вероятность пребывания формант в k -й полосе частот; $P_k(E_k)$ – коэффициент восприятия речи; E_k – эффективный уровень ощущения речевого сигнала в k -й полосе частот Δf_k .

Отличие методов заключается в том, что если в формантном методе используют тестовый сигнал $x(t)$ в виде стационарного случайного процесса (ССП), спектр мощности которого подобен долговременному спектру речи, а затем оценивают E_k в соответствии с соотношением:

$$E_k \approx \text{SNR}_k = 10 \lg \frac{D_{sk}}{D_{nk}}, \quad (2)$$

где SNR_k – парциальное отношение сигнал-шум;

D_{sk} и D_{nk} – соответственно дисперсии стационарных сигнала и шума в k -й полосе частот; то в ФМ методе для оценивания E_k предложено использовать иной способ, позволяющий корректно учесть влияние реверберационной помехи.

Это косвенный способ измерения отношения сигнал-шум, согласно которому используют тестовый сигнал $x(t)$ в виде нестационарного случайного процесса (НСП), дисперсия которого $D_x(t)$ модулирована по гармоническому закону с частотой модуляции F :

$$x(t) = \xi(t)\sqrt{f(t)}, \quad f(t) = 1 + \cos 2\pi Ft, \quad (3)$$

$$D_x(t) = D_\xi(1 + \cos 2\pi Ft), \quad (4)$$

где $\xi(t)$ – ССП с дисперсией D_ξ и спектром мощности, повторяющим форму долговременного спектра речи.

Дисперсия $D_y(t)$ НСП $y(t)$ будет также модулирована по гармоническому закону с той же частотой F , однако, вследствие влияния реверберационной и шумовой помех коэффициент модуляции дисперсии $D_y(t)$ оказывается меньше единицы. В полной версии ФМ метода при измерениях используют набор из 14 частот модуляции ($F_i = 0,63; 0,8; 1; 1,25; 1,6; 2; 2,5; 3,15; 4; 5; 6,3; 8; 10; 12,5$ Гц) и измеряют «эквивалентное» отношение сигнал-шум:

$$E_k = \frac{1}{14} \sum_{i=1}^{14} \text{SNR}_{ik}; \quad (5)$$

$$\text{SNR}_{ik} = 10 \lg \frac{m_k(F_i)}{1 - m_k(F_i)}, \quad (6)$$

где $m_k(F_i)$ – коэффициент модуляции дисперсии $D_y(t)$ сигнала $y(t)$, наблюдаемого в k -й полосе частот. Коэффициент модуляции $m_k(F_i)$ оценивают с использованием преобразования Фурье:

$$\tilde{m}_k(F_i) = \frac{2|A_{ik}(F_i)|}{|A_{ik}(0)|}, \quad A_{ik}(f) = \frac{1}{T} \int_0^T y_{ik}^2(t) e^{-j2\pi ft} dt, \quad (7)$$

где $|\cdot|$ – символ модуля; T – длительность реализации процесса

$$y_{ik}(t) = \int_{-\infty}^{\infty} h(v)x_{ik}(t-v)dv + n(t),$$

где $x_{ik}(t) = \xi_k(t)\sqrt{1 + \cos 2\pi F_i t}$ – модулированный полосовой белый шум в k -й полосе частот; $\xi_k(t)$ – результат фильтрации ССП $\xi(t)$ k -м полосовым фильтром.

В работе [11] предложена ускоренная версия ФМ метода, в которой вместо 14 измерительных опытов с тестовыми сигналами (3) используют единственный тестовый сигнал:

$$x(t) = \xi(t)\sqrt{f_5(t)}, \quad f_5(t) = 1 + 0,32 \sum_{i=1}^5 \sin 2\pi F_i t,$$

$$F_i = iF, \quad F = 0,7 \text{ Гц.}$$

При этом коэффициент модуляции оценивают в соответствии с соотношением

$$\tilde{m}_k(F_i) = \frac{5|A_{ik}(F_i)|}{|A_{ik}(0)|}, \quad (8)$$

а вместо соотношения (6) используют его аналог:

$$E_k = \frac{1}{5} \sum_{i=1}^5 \text{SNR}_{k,i}. \quad (9)$$

2. Оценивание коэффициента передачи модуляции для речевых тестовых сигналов

При использовании речевых тестовых сигналов можно использовать как «обычную», так и «быструю» версии ФМ метода, с тем лишь отличием, что коэффициент передачи модуляции $m_k(F_i)$ оценивают по результатам специальной обработки входного и выходного речевых сигналов.

В работе [8] предложено оценивать коэффициент передачи модуляции $m_k(F_i)$ в соответствии с выражением:

$$m_k(F_i) = \alpha \frac{\text{Re}\{S_{kXY}(F_i)\}}{S_{kXX}(F_i)}, \quad (10)$$

где $S_{kXY}(f)$ – взаимный спектр сигналов $z_X(t)$ и $z_Y(t)$ в k -й полосе частот; $z_X(t)$ и $z_Y(t)$ – огибающие интенсивностей речевых сигналов входного (тестового) сигнала $x(t)$ и выходного (в точке прослушивания) сигнала $y(t)$; $\text{Re}\{A\}$ – оператор выделения действительной части комплексного числа A ; $S_{kXX}(f)$ – спектр мощности входного сигнала; $\alpha = M\{z_X(t)\}/M\{z_Y(t)\}$ – нормализующий множитель; $M\{f(t)\}$ – оператор вычисления среднего значения функции $f(t)$.

Сигнал $z_X(t)$ (и аналогично – сигнал $z_Y(t)$) будем вычислять, исходя из модели полосового сигнала $x(t)$ в виде:

$$x(t) = X(t) \cos(2\pi f_0 t + \varphi(t)),$$

где $X(t)$ – огибающая сигнала $x(t)$; $\varphi(t)$ – фаза сигнала $x(t)$; f_0 – частота гармонической несущей.

Квадрат огибающей полосового сигнала можно оценивать в соответствии с выражением:

$$z_X(t) = \frac{1}{T_0} \int_{t-T_0}^t X^2(t) \cos^2(2\pi f_0 t + \varphi(t)) dt.$$

Действительно, при выполнении условия $T_0 \gg 1/f_0$ ($f_0 = 125, 250, 500, 1000, 2000, 4000$,

8000 Гц) справедливо приближенное выражение

$$z_X(t) \approx \frac{1}{2T_0} \int_{t-T_0}^t X^2(t) dt.$$

Кроме того, при выполнении условия $T_0 \ll T_1$, где $T_1 = 1/F_1$ – период модуляционной частоты из диапазона 0,63...12,5 Гц, справедливо

$$z_X(t) \approx \frac{X^2(t)}{2}.$$

Для экспериментальных исследований выберем значение параметра $T_0 = 0,01$ с, которое достаточно хорошо согласуется с рядом значений T_1 (0,08, ..., 1,5 с) и рядом значений $1/f_0$ (0,000125, ..., 0,008 с).

Поскольку функция $z_X(t)$ медленная по сравнению с функцией $x(t)$, можно существенно, почти на два порядка, снизить частоту ее дискретизации. Действительно, в работе [15] для спектра функции $z_X(t)$ принято разрешение по частоте $\Delta f = 0,24$ Гц, при этом параметр алгоритма БПФ принят равным $N_{fft} = 1024$. Нетрудно определить, что данные значения параметров соответствуют частоте дискретизации $F_s = 250$ Гц. Таким образом, при снижении частоты дискретизации с 22 050 Гц до 250 Гц выигрыш в объеме вычислений оказывается близким 88. Очевидно, после такого снижения частоты дискретизации спектральные выборки спектров $S_{kXY}(f)$ и $S_{kXX}(f)$ на модуляционных частотах от 0,63 Гц до 12,5 Гц – это 50 выборок с номерами от 3 до 52.

3. Результаты экспериментальных исследований

На рис. 1, 2 представлены результаты вычислений словесной разборчивости речи в условиях действия шумовой помехи для двух видов тестового сигнала: речевого (рис. 1) и шумового (рис. 2).

Сопоставляя приведенные графики, видим, что согласование достаточно хорошее на качественном уровне, однако количественно расхождения весьма заметны, достигая 30...40%.

На рис. 3, 4 представлены аналогичные результаты для условий совместного действия шума и реверберации (время реверберации $T_{30}, T_{60} \approx 2,9$ с). Сопоставляя приведенные графики, видим, что в условиях интенсивной реверберации наблюдаются существенные количественные и качественные расхождения.

Таким образом, полученные результаты свидетельствуют, с одной стороны, о принципиальной возможности использования речевого тестового сигнала в ФМ методе. С другой стороны, эти результаты показывают, что при использовании рече-

вого тестового сигнала погрешность измерений разборчивости речи может быть весьма высокой.

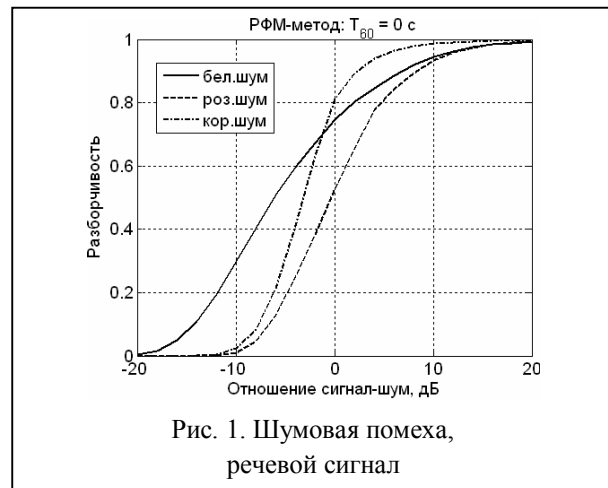


Рис. 1. Шумовая помеха, речевого сигнала

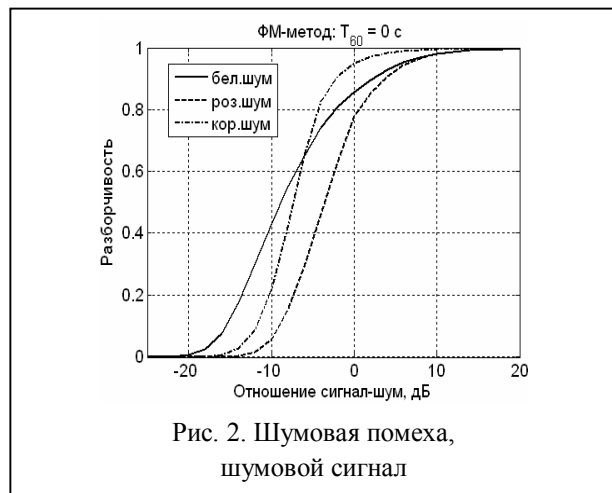


Рис. 2. Шумовая помеха, шумового сигнала

Следовательно, предложенный Друллманом [8] способ оценивания коэффициента передачи модуляции нуждается в коррекции. Один из способов такой коррекции предложен Гилсом [9], однако чрезмерная лаконичность описания этого способа препятствует его экспериментальной проверке.

Выводы

Произведена экспериментальная проверка способа оценивания коэффициента передачи модуляции, основанного на спектральном анализе интенсивностей огибающих тестового и выходного сигналов.

Показано, что данный способ обеспечивает получение правдоподобных результатов в условиях действия шумовой помехи, однако в условиях интенсивной реверберации данный способ практически не работоспособен. В этой связи целесообразно дальнейшие исследования направить на поиск способа оценивания коэффициента передачи модуляции, лишённого указанного недостатка.

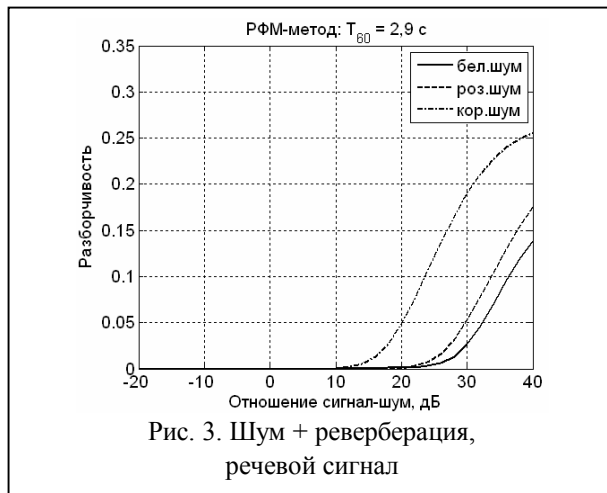


Рис. 3. Шум + реверберация, речевого сигнала

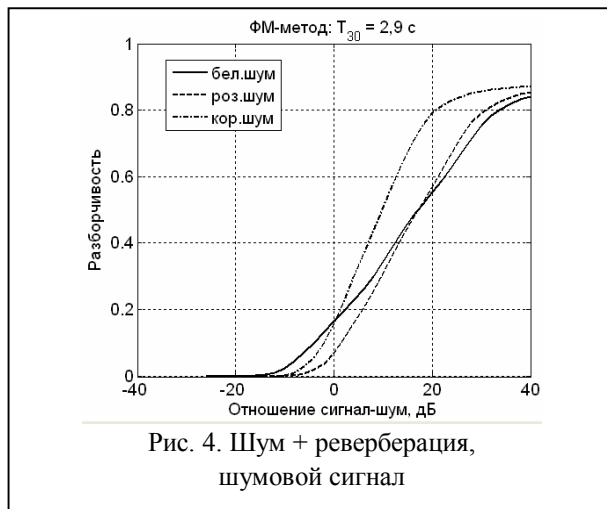


Рис. 4. Шум + реверберация, шумовой сигнала

Список литературы

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

Поступила в редколлегию 19.07.2013

Рецензент: д-р техн. наук, проф. А.И. Рыбин, Национальный технический университет Украины «КПИ», Киев.

МОВЛЕННЄВИЙ ТЕСТОВИЙ СИГНАЛ У ФОРМАНТНО-МОДУЛЯЦІЙНОМУ МЕТОДІ ОЦІНЮВАННЯ РОЗБІРЛИВОСТІ МОВЛЕННЯ

А.М. Продеус

Розглянуто принципову можливість вживання тестових сигналів у вигляді природного мовлення у формантно-модуляційному методі оцінювання розбірливості мовлення. Показано, що спосіб оцінювання коефіцієнта передачі модуляції, заснований на спектральному аналізі інтенсивностей огинаючих тестового і вихідного сигналів, забезпечує здобуття правдоподібних результатів в умовах дії шумової перешкоди, проте в умовах інтенсивної реверберації даний спосіб практично не є працездатним.

Ключові слова: оцінювання розбірливості мовлення, формантно-модуляційний метод, коефіцієнт передачі модуляції, шумова перешкода, реверберація.

USING OF SPEECH TEST SIGNAL IN FORMANT-MODULATION METHOD OF SPEECH INTELLIGIBILITY MEASURING

A.N. Prodeus

We consider the basic possibility of applying the test signals in the form of natural speech in formant-modulation method for estimating speech intelligibility. It is shown that the method of estimating a modulation transfer coefficient, based on the spectral analysis of envelope intensities of test and output signal, provides a plausible results under the action of noise, however, it is practically unworkable in an intensive reverberation conditions.

Keywords: evaluation of speech intelligibility, formant-modulation method, modulation transfer coefficient, noise interference, reverberation.