

---

УДК 621.391.83: 004.934.2

А.Н. Продеус, В.С. Дидковский, В.П. Овсяник

Национальный технический университет Украины «КПИ», Киев

## СЛЕПОЕ ИЗМЕРЕНИЕ ВРЕМЕНИ РЕВЕРБЕРАЦИИ В СИСТЕМАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

*При построении систем автоматического распознавания речи актуальной является задача коррекции речевых сигналов, искаженных реверберацией, для решения которой необходимо предварительно измерить время реверберации. Слепые измерения времени реверберации менее эффективны, в смысле точности распознавания речи, по сравнению с прямыми измерениями, однако в литературе трудно найти оценки такого проигрыша. В данной работе указанный пробел восполнен и указаны условия, выполнение которых позволяет минимизировать этот проигрыш, сведя его до 3-5%.*

**Ключевые слова:** измерение времени реверберации; слепые измерения; автоматическое распознавание речи; точность распознавания речи.

### Введение

Речевой сигнал в помещениях и линиях связи практически всегда подвергается искажающему действию реверберации, что отрицательно сказывается на эффективности систем автоматического распознавания речи (АРР) [1]. Поэтому задача построения систем АРР, робастных (устойчивых) к воздействию реверберации, весьма актуальна. Одним из способов ее решения является создание препроцессоров, корректирующих искаженные речевые сигналы до начала процедуры автоматического распознавания [2].

В работах [3, 4] показана возможность коррекции речевых сигналов путем подавления поздней

реверберации с использованием методов частотной коррекции и спектрального вычитания. При этом одним из необходимых условий возможности подавления поздней реверберации является наличие информации о времени реверберации  $T_{60}$  канала передачи. Обычно время реверберации  $T_{60}$  оценивают «напрямую», используя импульсную характеристику  $h(t)$  канала передачи [5]. Если импульсная характеристика неизвестна, приходится прибегать к так называемым «слепым» измерениям, в соответствии с которыми информацию о времени реверберации  $T_6$  извлекают из реверберированного речевого сигнала [6, 7].

Результаты исследований свидетельствуют, что слепые измерения  $T_6$  менее точны, нежели прямые

измерения. Другим недостатком слепых измерений является приближенный характер оценивания частотной зависимости параметра  $T_{60}$  [8, 9]. Оба упомянутых фактора негативно сказываются на качестве функционирования системы АРР, снижая точность распознавания речи. К сожалению, степень такого снижения до последнего времени оставалась неизвестной, что в значительной степени поясняется как сложностью теоретического аспекта данной проблемы, так и сложностью построения соответствующих экспериментальных программных комплексов. Цель данной работы состояла в преодолении указанных трудностей путем проведения экспериментальных исследований с использованием программного комплекса, построенного на базе современных алгоритмов оценивания спектра поздней реверберации и слепого измерения времени реверберации.

## 1. Постановка задачи

Модель речевого сигнала  $y(t)$ , искаженного реверберацией, можно представить в виде свертки «чистого» речевого сигнала  $x(t)$  с импульсной характеристикой  $h(t)$  помещения:

$$y(t) = \int_0^{\infty} h(v)x(t-v)dv = x(t) \otimes h(t),$$

где  $\otimes$  – символ свертки.

Используя модель Полэка [1] для импульсной характеристики (ИХ) помещения

$$h(t) = \xi(t)e^{-\delta t},$$

где  $\xi(t)$  – стационарный белый шум,  $\delta = 3 \ln 10 / T_{60}$  – показатель скорости затухания уровня звука в помещении, а также выделяя в ИХ  $h(t)$  (рис. 1) области, соответствующие ранним и поздним отражениям:

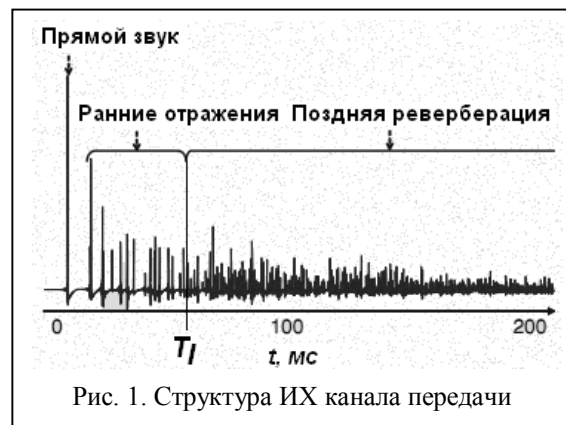
$$h_i(t) = \begin{cases} h(t), & 0 \leq t \leq T_1; \\ 0, & \text{др. } t \end{cases}$$

$$h_l(t) = \begin{cases} h(t+T_1), & t \geq 0; \\ 0, & \text{др. } t \end{cases},$$

искаженный реверберацией сигнал можно представить в виде:

$$y(t) = h_i(t) \otimes x(t) + r(t), \quad (1)$$

где  $r(t) = h_l(t) \otimes x(t - T_1)$  – компонент, обусловленный действием поздней реверберации;  $T_1$  – момент начала поздней реверберации (рис. 1). Полагая, что слагаемые соотношения (1) статистически независимы, приходим к трактовке поздней реверберации как разновидности аддитивного шума. К сожалению, нестационарный характер этого шума делает неэффективным непосредственное применение традиционных методов подавления шума, рассчитанных на стационарный или медленный нестационарный шум.



Традиционными методами подавления шумовой помехи являются метод частотной коррекции [10] и метод спектрального вычитания, согласно которым оценку спектра мощности 1-го фрейма сигнала  $x(t)$  для  $k$ -й частотной выборки, можно представить в виде:

$$\hat{\lambda}_x^{1/2}(l, k) = G(l, k)\lambda_y^{1/2}(l, k),$$

где  $\lambda_y(l, k)$  – спектр мощности 1-го сегмента сигнала  $y(t)$  на частоте  $f_k = kF_s / N_{\text{fft}}$ ;  $F_s$  – частота дискретизации;  $N_{\text{fft}}$  – параметр быстрого преобразования Фурье;  $k$  – номер частотной выборки;  $G(l, k)$  – коэффициент передачи корректирующего фильтра для 1-го фрейма сигнала  $y(t)$  на частоте  $f_k$ . Фазовый спектр восстановленного сигнала  $\hat{x}(t)$  обычно принимают равным фазовому спектру искаженного сигнала  $y(t)$ .

Поскольку поздняя реверберация, согласно (1), является разновидностью аддитивного шума, метод частотной коррекции несложно модернизировать под задачу подавления поздней реверберации, заменив оценку спектра шумовой помехи  $\lambda_n(l, k)$  оценкой спектра поздней реверберации  $\lambda_r(l, k)$ :

$$\lambda_r(l, k) = e^{-2\delta(k)T_1} \cdot \lambda_y(l - N_1, k), \quad (2)$$

где  $N_1 = T_1 F_s / R$ ;  $R$  – сдвиг фреймов, выраженный в выборках. Заметим, что соотношение (2) справедливо для расстояний между источником звука и микрофоном, больших критической дистанции (в данной работе ограничимся рассмотрением этой ситуации) [1].

Для повышения точности измерений спектра  $\lambda_y(l, k)$  в [3] предложено производить скользящее усреднение:

$$\hat{\lambda}_y(l, k) = \eta_z \hat{\lambda}_y(l-1, k) + (1 - \eta_z) |Y(l, k)|^2, \quad (3)$$

где  $\eta_z$  – параметр, регулирующий степень усреднения ( $0 \leq \eta_z < 1$ );  $Y(l, k)$  – дискретное преобразование Фурье 1-го фрейма сигнала  $y(t)$ .

В работе [4] предложен более сложный алгоритм усреднения:

$$\hat{\lambda}_y(l, k) = \eta_z(k) \hat{\lambda}_y(l-1, k) + (1 - \eta_z(k)) |Y(l, k)|^2 \quad (4)$$

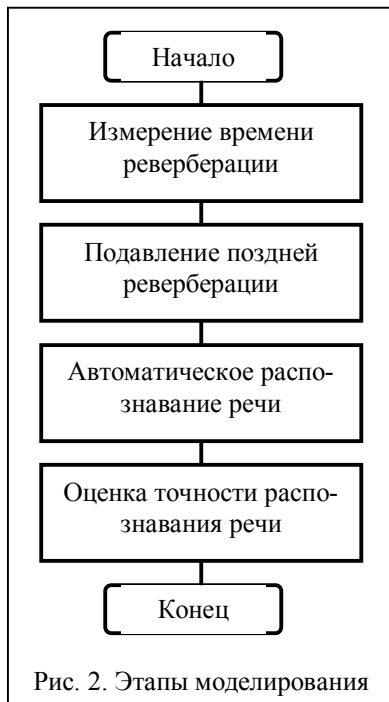
$$\eta_z(k) = \begin{cases} \eta_z^d(k), & |Y(l, k)|^2 \leq \hat{\lambda}_y(l-1, k); \\ \eta_z^a(k), & \text{в остальных случаях,} \end{cases} \quad (5)$$

$$0 \leq \eta_z^a(k) < \eta_z^d(k). \quad (6)$$

При этом верхнее значение параметра  $\eta_z^d(k)$  предложено ограничивать величиной

$$\eta_z^d(k) = \frac{1}{1 + 2\delta(k)R/F_s}, \quad \delta(k) = \frac{3 \ln 10}{T_{60}(k)}. \quad (7)$$

Из соотношения (2) следует, что подавление поздней реверберации возможно при условии, что известно время реверберации  $T_{60}$ , которое обычно измеряют либо по имеющейся импульсной характеристике  $h(t)$  канала передачи, либо так называемыми слепыми методами, т.е. по речевым сигналам, искаженным реверберацией. В данной работе, с использованием методов компьютерного моделирования, производится сопоставление эффективности обоих подходов с использованием такого критерия как точность автоматического распознавания откорректированной речи. Этапы такого моделирования представлены на рис. 2.



## 2. Измерение времени реверберации методом максимального правдоподобия

Если ИХ  $h(t)$  канала передачи известна, измерение времени реверберации  $T_{60}$  обычно производят с использованием хорошо известного метода

Шредера [5]. Что касается слепого измерения времени реверберации, здесь существует многообразие методов. В работах [11] и [12] изложен нейросетевой подход. Подход, базирующийся на таком сегментировании речи, при котором выделяются паузы, предложен в [3]. В работах [13] и [14] представлен подход, при котором предполагается, что частотная характеристика фильтра, обратного ИХ помещения, является минимально-фазовой (что на практике обычно не выполняется).

В данной работе использован иной метод, впервые предложенный в работе [6] и названный методом максимального правдоподобия (МП). Основанием для такого выбора является хорошая помехоустойчивость метода МП, не уступающего (и даже несколько превосходящего) конкурентным методам в точности измерений [15].

Кроме того, организацию и проведение соответствующих экспериментальных исследований существенно облегчает наличие в открытом доступе программы Matlab для слепого оценивания времени реверберации методом МП [16].

Изложим вкратце сущность измерений времени реверберации  $T_{60}$  методом МП. Информацию о параметре  $T_{60}$  извлекают преимущественно в паузах речевого сигнала, где действие реверберации проявляется в виде звуковых «шлейфов», тянущихся за последними звуками слов.

Структура этих шлейфов подобна структуре ИХ канала передачи:

$$y(n) = \xi(n)a(n),$$

$$a(n) = e^{-\delta n/F_s}$$

где  $\xi(n)$ ,  $n \geq 0$ ;  $n = tF_s$  – дискретный гауссовский белый шум с параметрами  $[0, \sigma]$ .

Если наблюдается  $N$ -мерный вектор  $\mathbf{y}$ , тогда в силу статистической независимости выборок процесса  $\xi(n)$ , многомерная плотность вероятностей этого вектора, именуемая также функцией правдоподобия, имеет вид произведения  $N$  одномерных гауссовских распределений:

$$L(\mathbf{y}; \mathbf{a}, \sigma) = \frac{1}{a(0) \cdots a(N-1)} \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \times \exp \left( - \frac{\sum_{n=0}^{N-1} (y(n)/a(n))^2}{2\sigma^2} \right). \quad (8)$$

Используя соотношение (8), необходимо оценить параметры  $\sigma$  и  $\mathbf{a}$ , где  $\mathbf{a}$  –  $N$ -мерный вектор, по имеющимся  $N$  значениям вектора  $\mathbf{y}$ . Учитывая экспоненциальный характер функции  $a(n)$ :

$$a(n) = a^n, \quad a = \exp(-\delta/F_s), \quad (9)$$

из (8) с учетом (9) следует

$$L(y; a, \sigma) = \left( \frac{1}{2\pi a^{(N-1)} \sigma^2} \right)^{N/2} \times \exp \left( - \frac{\sum_{n=0}^{N-1} a^{-2n} y(n)^2}{2\sigma^2} \right). \quad (10)$$

Логарифмируя (10), получим

$$\ln L(y; a, \sigma) = - \frac{N(N-1)}{2} \ln(a) - \frac{N}{2} \ln(2\pi\sigma^2) - \frac{\sum_{n=0}^{N-1} a^{-2n} y(n)^2}{2\sigma^2}. \quad (11)$$

Приравнявая нулю частные производные от выражения (11) по параметрам  $\sigma$  и  $a$ , получаем систему уравнений для неизвестных  $\sigma$  и  $a$ :

$$\frac{\sum_{n=0}^{N-1} na^{-2n} y(n)^2}{a\sigma^2} - \frac{N(N-1)}{2a} = 0, \quad (12)$$

$$\sigma^2 = \frac{1}{N} \sum_{n=0}^{N-1} a^{-2n} y(n)^2. \quad (13)$$

Подставляя (13) в (12), получаем соотношение, определяющее параметр  $a$ :

$$\frac{\sum_{n=0}^{N-1} na^{-2n} y(n)^2}{\sum_{n=0}^{N-1} a^{-2n} y(n)^2} - \frac{(N-1)}{2} = 0. \quad (14)$$

В силу нелинейности уравнения (14) алгоритм нахождения неизвестного параметра  $a$  должен быть итерационным.

Формируя таким образом оценку  $a_k^*$  в  $k$ -м фрейме речевого сигнала, для множества фреймов получаем множество оценок, по которым можно построить гистограмму. Часть оценок хороши, поскольку соответствуют фреймам, захватывающим паузы, а часть оценок неудачны, т.к. захватывают речевой сигнал. Кроме того, различные участки пауз обладают неодинаковой информативностью. Очевидно, предпочтительными являются те участки пауз, которые примыкают к окончаниям слов. Поэтому необходимо решающее правило для выбора информативных участков пауз, а также для выбора хорошей оценки параметра  $a$  с использованием гистограммы.

Такое решающее правило может базироваться на следующих рассуждениях. Если внутри текущего фрейма, в пределах двух смежных участков времени (субфреймов) энергия сигнала снижается, тогда принимают решение о наличии участка паузы, на котором замечено действие реверберации. Субфреймы объединяют в так называемый «сегмент данных», по которому вычисляют «текущее» время реверберации  $T_{60}$ , решая уравнение (14) и используя соотношения

$$T_{60} = 6,91/\delta, \quad \delta = -F_s \ln a. \quad (16)$$

После этого обновляют гистограмму оценок  $T_{60}$ , сформированную из последних оценок  $T_{60}$ .

Оценку  $\hat{T}_{60}^{(1)}$ , соответствующую максимуму обновленной гистограммы, считают текущей оценкой времени реверберации. Для снижения дисперсии оценки времени реверберации используют процедуру сглаживания:

$$\hat{T}_{60}(\lambda) = \beta(\lambda) \cdot \hat{T}_{60}(\lambda-1) + (1-\beta(\lambda)) \cdot \hat{T}_{60}^{(1)}(\lambda), \quad (17)$$

где  $0 < \beta(\lambda) < 1$ ;  $\lambda$  - номер фрейма.

### 3. Организация экспериментальных исследований

Реверберированные сигналы моделировались в среде Matlab путем свертки «чистого» (записанного при отношении сигнал-шум не менее 40 дБ) речевого сигнала с ИХ трех помещений с временами реверберации 0,74 с, 0,89 с и 1,1 с. Параметры оцифровки звуковых файлов: частота дискретизации 22050 Гц, равномерное квантование 16 бит.

Слепое измерение времени реверберации производилось в среде Matlab с использованием пакета программ ML\_RT\_estimation, реализующего измерение параметра  $T_{60}$  методом МП [16]. Подпрограмма-функция ML\_RT\_estimation\_init.m этого пакета подверглась незначительной модернизации: значение параметра  $\beta(\lambda)$ , используемого в соотношении (17), было снижено с 0,996 до 0,9, а значение начального времени  $\hat{T}_{60}^{(1)}$  было повышено с 0,3 с до 0,5 с. Такая модернизация позволила ускорить сходимость процедуры оценивания.

Система подавления реверберации также моделировалась в Matlab, с использованием модифицированной, в соответствии с соотношениями (2) – (7), программы-функции ssubmmse.m, входящей в состав программного инструментария VoiceBox [17]. При этом фреймы речевого сигнала, протяженностью 32 мс, взвешивались окном Хэмминга и перекрывались на 50%.

Коэффициент передачи корректирующего фильтра вычислялся в соответствии с алгоритмом logMMSE [18]:

$$G(l, k) = \frac{\xi(l, k)}{1 + \xi(l, k)} \exp \left( \frac{1}{2} \int_{v(l, k)}^{\infty} \frac{e^{-t}}{t} dt \right),$$

$$v(l, k) = \frac{\xi(l, k)}{1 + \xi(l, k)} \gamma(l, k).$$

Результаты экспериментов оценивались качественно и количественно. Качественная оценка результатов дереверберации состояла в прослушивании звуковых файлов реверберированного и дереверберированного сигналов, а также в визуальном анализе соответствующих спектрограмм.

Количественная оценка результатов дереверберации состояла в вычислении значений показателя Acc% [9], именуемого «точностью правильного распознавания слов»:

$$Acc\% = \frac{N - D - S - I}{N} \times 100\% ,$$

где N – общее количество распознаваемых слов; D – количество ошибочных удалений слов; S – количество ошибочно замененных слов; I – количество ошибочно вставленных слов.

Важным параметром алгоритма дереверберации является  $T_1$  – граница между ранними отражениями и поздней реверберацией. В работе [19] было показано, что оптимальным, в смысле максимума Acc%, является значение  $T_1$ , принадлежащее интервалу 100-200 мс. В данной работе принято  $T_1 = 98$  мс. Кроме того, в [19] была продемонстрирована затруднительность получения аналогичных значений  $T_1$ , оптимальных в смысле максимума показателя PESQ [20], характеризующего качество речевого сигнала. Поскольку данное обстоятельство, по существу, означает непригодность показателя PESQ для оценки качества деревербератора в системе APP, оценивание PESQ в данной работе носило преимущественно иллюстративный характер.

Для моделирования системы APP, а также для оценки показателя Acc%, применялся программный инструментальный НТК [21]. Обучение системы APP производилось с использованием 269 образцов 27 слов украинской речи, произнесенных 2-мя дикторами женщинами. Фонемный словарь состоял из 27 элементов, использовались 39-мерные классификационные признаки вида MFCC\_0\_D\_A. Тестовый сигнал представлял собой звуковой файл дискретной речи с записью поочередно зачитываемых всех 27 слов, использованных при обучении. Паузы между словами при этом составляли 0,2...0,5 с.

#### 4. Результаты экспериментов

Было проведено две группы экспериментов, в каждой из которых сравнивались результаты дереверберации, осуществленные с использованием оценок  $T_{60}$ , полученных прямым и слепым способами. В первой группе экспериментов слепое измерение времени реверберации производилось в предположении, что параметр  $T_{60}$  не зависит от частоты. Во второй группе экспериментов при слепом измерении времени реверберации приближенно оценивалась зависимость  $T_{60}(f)$ .

##### Первая группа экспериментов

На слух результаты дереверберации, полученные путем измерения  $T_{60}$  по ИХ помещения, прак-

тически не отличаются от таковых, где  $T_{60}$  оценивалось вслепую.

На рис. 3 приведены спектрограммы исходного (а) и реверберированного (б) сигналов, а также сигналов, подвергнутых процедуре дереверберации: для  $T_{60}$ , измерявшегося по ИХ помещения (в) и для  $T_{60}$ , измерявшегося вслепую (г). Хорошо видно, что реверберация действует на сигнал как НЧ фильтр, подавляя высокочастотные составляющие согласных звуков. Вместе с тем видно, что после дереверберации затянutosть сигнала по времени действительно уменьшилась, хотя мощность высокочастотных компонентов так и не восстановилась полностью.

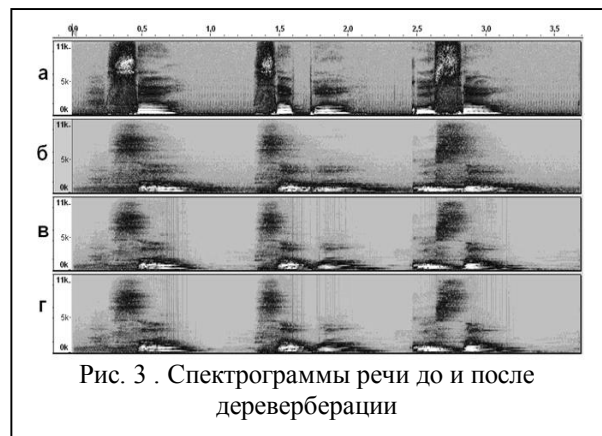


Рис. 3 . Спектрограммы речи до и после дереверберации

Результаты слепого оценивания параметра  $T_{60}$  приведены в табл. 1.

Таблица 1  
Результаты слепого оценивания  $T_{60}$

Оценка по ИХ $T_{60}$ , с	Слепая оценка $T_{60}$ , с	Абсолют. погрешн., с	Относит. погрешн., %
0,74	0,65	0,09	13,8
0,89	0,7	0,19	21,3
1,1	0,8	0,3	27,3

Как видим, абсолютная и относительная погрешности слепой оценки составили 0,1-0,3 с и 14%-27%, соответственно.

Результаты оценивания точности распознавания Acc% сведены в табл. 2 и представлены графиком рис. 4, где обозначено:

«способ 1» – попытка коррекции реверберированной речи с помощью традиционного алгоритма подавления шумовой помехи;

«способ 2» – коррекция путем подавления поздней реверберации при оценивании  $T_{60}$  по ИХ помещения;

«способ 3» - коррекция путем подавления поздней реверберации при слепом оценивании  $T_{60}$ .

Таблица 2

Асс% в отсутствие и при наличии коррекции

T <sub>60</sub> , с	Точность распознавания, Асс%			
	нет корр-ции	корр-ция спос. 1	корр-ция спос. 2	корр-ция спос. 3
0,74	22,22	18,52	74,1	37,04
0,89	22,22	14,81	55,6	37,04
1,1	29,63	29,63	62,3	62,96
<b>Сред.</b>	<b>24,69</b>	<b>21,00</b>	<b>64,00</b>	<b>45,68</b>



Рис. 4. Точность распознавания в отсутствие и при наличии коррекции речевого сигнала

Как видим, использование усредненной оценки T<sub>60</sub>, измеренной вслепую (способ 3), существенно ухудшило качество распознавания для помещений с временем реверберации 0,74 с и 0,89 с.

Исключением явилось помещение с временем реверберации 1,1 с.

В этом случае точность распознавания не ухудшилась, а даже несколько повысилась. Тем не менее, в среднем слепое измерение T<sub>60</sub> приводит к снижению Асс% примерно на 20%.

Как следует из табл. 3 и рис. 5, качество речи, характеризуемое показателем PESQ, при слепом оценивании T<sub>60</sub> также несколько ухудшилось по сравнению с ситуацией, когда T<sub>60</sub> оценивалось по ИХ помещения.

Таблица 3

PESQ в отсутствие и при наличии коррекции

T <sub>60</sub> , с	PESQ			
	нет корр.	корр. спос. 1	корр. спос. 2	корр. спос. 3
0,74	2,281	2,252	2,33	2,308
0,89	2,073	2,059	2,08	2,111
1,1	2,03	2,037	2,23	2,212
<b>Сред.</b>	<b>2,128</b>	<b>2,116</b>	<b>2,213</b>	<b>2,210</b>

Здесь также не обошлось без исключения – для помещения с временем реверберации 0,89 с качест-

во речи повысилось. Однако в среднем качество речи осталось практически неизменным, что свидетельствует о практической непригодности показателя PESQ для оценки качества деревьевербераторов в системах APP.

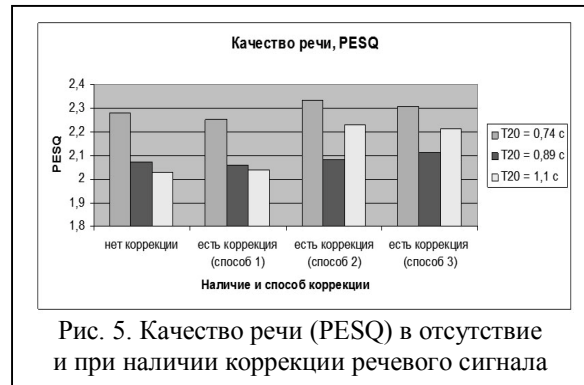


Рис. 5. Качество речи (PESQ) в отсутствие и при наличии коррекции речевого сигнала

### Вторая группа экспериментов

Во второй группе экспериментов при слепом измерении времени реверберации оценивалась зависимость T<sub>60</sub>(f). С этой целью реверберированный речевой сигнал подвергался фильтрации гребенкой фильтров. Время реверберации оценивалось вслепую по речевому сигналу на выходе каждого из фильтров. Затем по полученным результатам, путем интерполяции и экстраполяции, формировалась зависимость T<sub>60</sub>(f). Результаты прямого оценивания T<sub>60</sub>(f), осуществленного по спектрограмме имеющейся ИХ, показаны на рис. 6 – 8 тонкой сплошной линией.

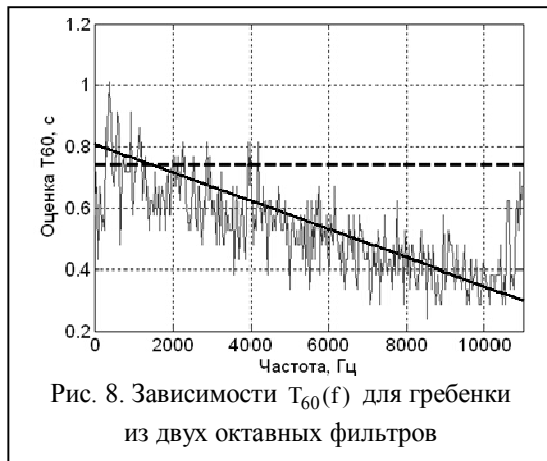
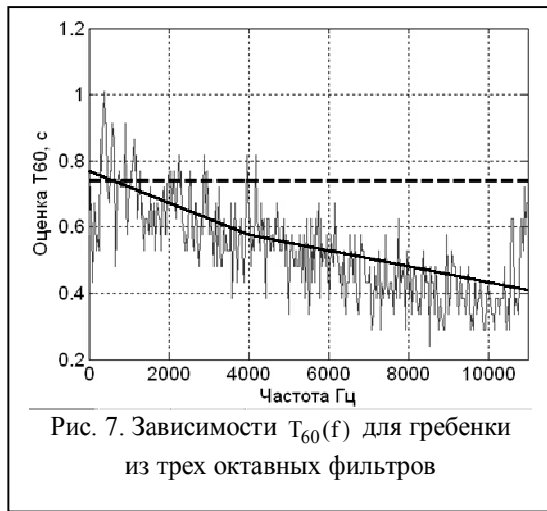
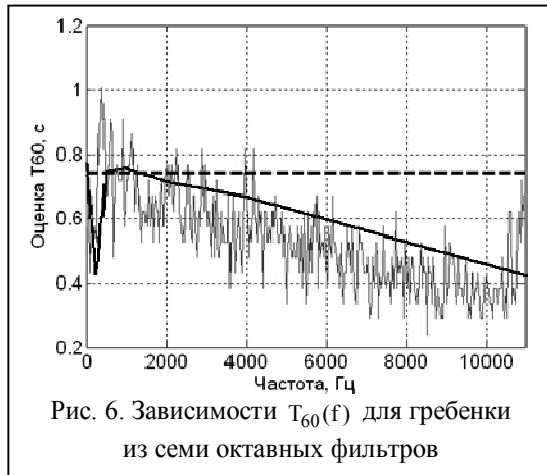
Результаты слепого оценивания T<sub>60</sub>(f) представлены на этих же рисунках жирной сплошной линией. Пунктирной линией показаны результаты прямого измерения T<sub>60</sub> без учета зависимости времени реверберации от частоты.

Результаты рис. 6 соответствуют гребенке из семи октавных фильтров с центральными частотами f<sub>0</sub> = 125, ..., 8000 Гц, рис. 7 – гребенке из трех октавных фильтров с центральными частотами 2 кГц, 4 кГц и 8 кГц, рис. 8 – гребенке из двух октавных фильтров с центральными частотами 2 и 8 кГц.

Практически те же результаты получаются при использовании гребенки третьооктавных фильтров.

Результаты оценивания точности распознавания Асс% во второй группе экспериментов сведены в табл. 4 и представлены графиком рис. 9. Нетрудно видеть, что гребенка из двух фильтров обеспечивает вполне приемлемое качество слепого оценивания зависимости T<sub>60</sub>(f).

При этом учет зависимости времени реверберации от частоты позволил повысить точность распознавания речи в среднем на 15-17% по сравнению со случаем, когда такой учет не производился.



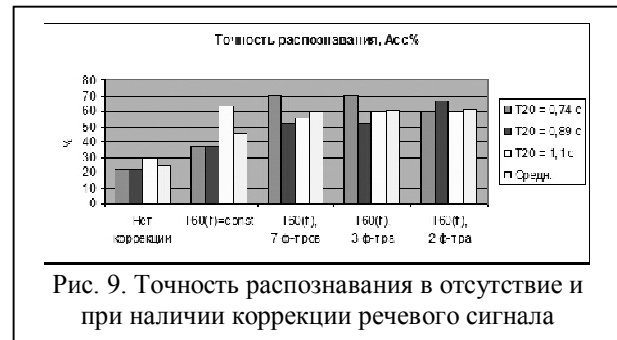
Сравнивая между собой полученные результаты, заключаем, что при слепом измерении времени реверберации можно достичь точности распознавания речи, всего лишь на 3 – 5% уступающей таковой для прямого измерения времени реверберации, если при слепом измерении учитывать зависимость времени реверберации от частоты. Достаточная для приложений точность измерений зависимости  $T_{60}(f)$  была достигнута при использовании гребенки, состоящей всего из двух октавных или треть-

октавных фильтров с центральными частотами 2 кГц и 8 кГц.

Таблица 4

Точность распознавания во второй группе экспериментов

$T_{60}$ , с	Точность распознавания, Асс%				
	Нет корр.	$T_{60}(f) = \text{const}$	$T_{60}(f)$ , 7 ф-в	$T_{60}(f)$ , 3 ф-тра	$T_{60}(f)$ , 2 ф-тра
0,74	22,22	37,04	70,37	70,37	59,26
0,89	22,22	37,04	51,85	51,85	66,67
1,1	29,63	62,96	55,56	59,26	59,26
<b>Сред.</b>	<b>24,69</b>	<b>45,68</b>	<b>59,26</b>	<b>60,49</b>	<b>61,73</b>



### Выводы

В данной работе, с использованием методов компьютерного моделирования, сопоставлены слепой и прямой способы измерения времени реверберации. Поскольку оценки времени реверберации предназначены для использования в деревербераторе, осуществляющем коррекцию речевого сигнала на входе системы АРР, в качестве меры эффективности способов измерения времени реверберации использован сквозной показатель в виде точности автоматического распознавания откорректированной речи.

Показано, что предположение о частотной независимости времени реверберации при слепых измерениях  $T_{60}$  методом максимального правдоподобия приводит к значительному снижению точности распознавания, по сравнению с таковой для прямых измерений. В рассмотренных примерах такой проигрыш оказался близким 20%.

Учет зависимости времени реверберации от частоты при слепых измерениях позволил снизить величину проигрыша до 3-5%. При этом достаточным оказалось использование гребенки, состоящей всего из двух октавных или третьоктавных фильтров, центральные частоты которых близки к границам частотного диапазона анализируемого речевого сигнала.

Вместе с тем, проведенные исследования показали неэффективность параметра PESQ как меры эффективности деревербератора, используемого в качестве препроцессора системы АРР.

## Список літератури

1. Naylor P. *Speech Dereverberation* / P. Naylor, N. Gaubitch – Springer-Verlag: London, 2010.
2. Yoshioka T. *Making Mashine Understand Us in Reverberant Rooms* / T. Yoshioka, A. Sehr, M. Delcroix et al. // *IEEE Signal Processing Magazine*. – 2012, Nov. – Vol. 29, No. 6. – P. 114-126.
3. Lebart K. *A new method based on spectral subtraction for speech dereverberation* / K. Lebart, J. Boucher, P. Denbigh // *Acta Acoustica*. – 2001. – Vol. 87, No. 3. – P. 359-366.
4. Habets E. *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement: dissert. ... Doctor of Philosophy Electrical Engineering / Emanuel Habets*. – Technische Universiteit Eindhoven, 2007. – 241 p.
5. Schroeder R.M. *New Method of Measuring Reverberation Time* / R.M. Schroeder. – *JASA*, 1964. – P. 409-412.
6. *Blind estimation of reverberation time* / Ratnam R., Jones D.L., Wheeler B.C. et al. // *J. Acoust. Soc. Am.* – 2003. – Vol. 114, No. 5. – P. 2877-2892.
7. *An improved algorithm for blind reverberation time estimation* // H. Lollmann, E. Yilmaz, M. Jeub, P. Vary // *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*. – 2010, Aug.
8. Lollmann H. *Estimation of the Frequency Dependent Reverberation Time by Means of Warped Filter-Banks* / H. Lollmann, P. Vary // *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. – 2011. – P.309-312.
9. Jeub M. *Joint Dereverberation and Noise Reduction for Binaural Hearing Aids and Mobile Phones: Master of Sc. Degree Dissert.* / M. Jeub. – Aachen, Germany, 2012. – 174 p.
10. *Noise Reduction in Speech Processing* / J. Benesty, J. Chen, Y. Huang, I. Cohen. – Springer-Verlag: Berlin, Heidelberg, 2009.
11. Nannariello J. *The prediction of reverberation time using neural network analysis* // J. Nannariello, F. Fricke // *Appl. Acoust.* – 1999. – No.58. – P. 305-325.
12. Cox T.J. *Extracting room reverberation time from speech using artificial neural networks* / T.J. Cox, F. Li, P. Darlington // *J. Audio Eng. Soc.* – 2001. – No.49. – P. 219-230.
13. Neely S.T. *Invertibility of room impulse response* / S.T. Neely, J.B. Allen // *J. Acoust. Soc. Am.* – 1979. – Vol.66. – P. 165-169.
14. Miyoshi M. *Inverse filtering of room impulse response* / M. Miyoshi, Y. Kaneda // *IEEE Trans. Acoust., Speech, Signal Process.* – 1988. – Vol.36. – P. 145-152.
15. Gaubitch N. *Performance comparison of algorithms for blind reverberation time estimation from speech* / N. Gaubitch, H.W. Lollmann, M. Jeub et al. // *Acoustic Signal Enhancement; Proc. of IWAENC 2012; Internat. Workshop on 4-6 Sept. 2012*. – P.1-4.
16. Jeub M. *Blind Reverberation Time Estimation* / M. Jeub [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/35740-blind-reverberation-time-estimation> (02.07.2014).
17. Brooks M. *VOICEBOX: Speech Processing Toolbox for MATLAB* / M. Brooks // Imperial College London, Electrical Engineering Department. – [Online]. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html> (02.07.2014).
18. Ephraim Y. *Speech enhancement using a minimum mean square error Log-spectral amplitude estimator* / Y. Ephraim, D. Malah // *IEEE Trans. Acoust., Speech, Signal Processing*. – 1985, Apr. – Vol. ASSP-33. – P. 443-445.
19. Prodeus A. *On existence of optimal boundary value between early reflections and late reverberation* / A. Prodeus, O. Ladoshko // *Electronics and Nanotechnology: IEEE 34th International Scientific Conference, 15-18 April 2014, Kyiv, Ukraine*. – Kyiv, 2014.
20. Beerends J. *Measurement of speech intelligibility based on the PESQ approach* / J. Beerends, E. Larsen, N. Iyer, J. van Vugt // *Measurement of Speech and Audio Quality in Networks (MESAQIN): Int. Conf., 2 June 2004, Prague, Czech Republic*. – Prague, 2004.
21. *The HTK Book* / S. Young, G. Evermann, M. Gales, et al. – Cambridge: University Engineering Department, 2009. – 375 p.

Поступила в редколлегию 8.07.2014

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

### ОЦІНКА ЕФЕКТИВНОСТІ ОДНОГО МЕТОДА СЛІПОГО ВИМІРЮВАННЯ ЧАСУ РЕВЕРБЕРАЦІЇ

А.М. Продеус, В.С. Дідковський, В.П. Овсяник

При побудові систем автоматичного розпізнавання мовлення актуальним є завдання коригування мовленнєвих сигналів, спотворених реверберацією, для розв'язання якого необхідно попередньо виміряти час реверберації. Сліпі вимірювання часу реверберації є менш ефективними, в сенсі точності розпізнавання мовлення, порівняно із прямими вимірюваннями, однак у літературі важко знайти оцінки такого програшу. У даній роботі зазначений недолік усунуто та вказано, за яких умов цей програш можна мінімізувати, звівши його до 3-5%.

**Ключові слова:** вимірювання часу реверберації; сліпе вимірювання; автоматичне розпізнавання мовлення; точність розпізнавання мовлення.

### PERFORMANCE EVALUATION OF A METHOD OF REVERBERATION TIME BLIND MEASUREMENTS

A.N. Prodeus, V.S. Didkovskiy, V.P. Ovsianyk

When building a system of automatic speech recognition, there is problem of reverberated speech correcting, for which solution it is necessary first measure a reverberation time. Blind measurements of reverberation time are less efficient, in the sense of speech recognition accuracy, compared with direct measurements, but it is hard to find in the literature estimations of this loss. In this paper, this gap is replenished and conditions, the implementation of which helps to minimize this loss by reducing it to 3-5%, are defined.

**Keywords:** reverberation time measurement; blind measurement; automatic speech recognition; speech recognition accuracy.