

Теория сигналов и систем

УДК 004.934

Программный инструментарий для тестирования систем обработки речевых сигналов. Часть 2. Оценка эффективности

А.Н. Продеус, д-р техн. наук

Национальный технический университет Украины «Киевский политехнический институт»,
пр. Победы, 37, Киев-56, 03056, Украина.

Предложен состав программного инструментария для исследований и оптимизации алгоритмов систем коррекции и кодирования речевых сигналов, подверженных различного рода искажениям, таким как шумовая и реверберационная помехи, ошибки кодирования. В первой части данной работы рассмотрен инструментарий, позволяющий решать такие задачи как создание корпусов зашумленной речи и коррекция зашумленных речевых сигналов. В данной, второй части работы, предложен программный инструментарий для аттестации тестируемых систем обработки речевых сигналов по таким критериям как качество и разборчивость речи, достоверность автоматического распознавания речи. Разработанный инструментарий является неоднородным: часть вычислений производится в среде Matlab, а часть – вне этой среды. Несмотря на указанную особенность, данный инструментарий позволяет совместно использовать все необходимые программные компоненты, что обеспечивает сквозную автоматизацию обработки речевых сигналов, начиная от их намеренного искажения, последующей коррекции – и заканчивая оценкой качества

восстановленных речевых сигналов. Библ. 6, рис. 2.

Ключевые слова: программный инструментарий, критерии качества, тестирование систем.

Введение

В первой части данной работы была сформулирована задача расширения состава и возможностей существующего программного инструментария, что позволило бы решать такие сложные задачи комплексного характера как сопоставление и оптимизация алгоритмов шумоподавления, входящих в состав математического обеспечения систем цифровой связи и систем автоматического распознавания речи.

В соответствии с обобщенной схемой тестирования систем обработки речевых сигналов (рис. 1), такой программный инструментарий должен состоять из трех компонентов:

- программы создания искаженных речевых сигналов;
- программы моделирования тестируемых систем;
- программы оценивания качества тестируемых систем.

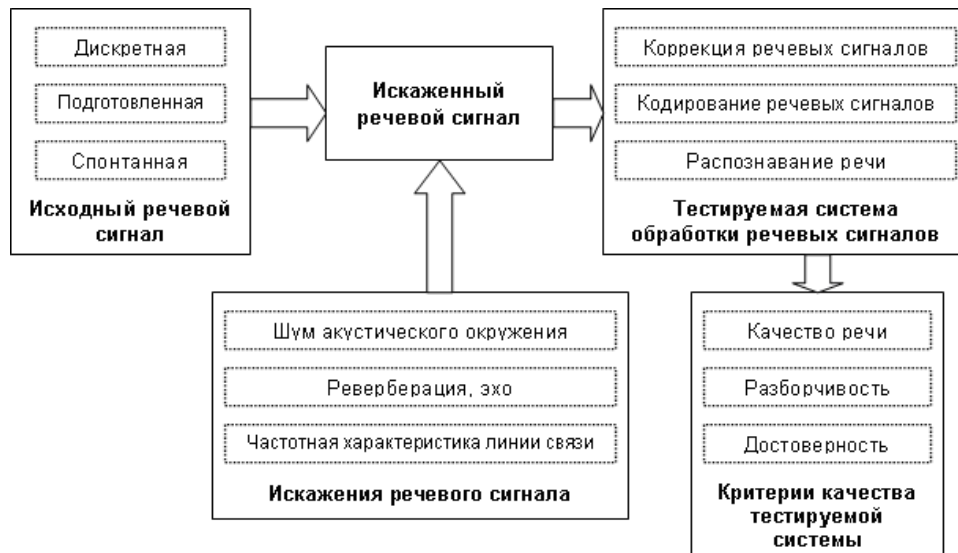


Рис. 1. Схема тестирования систем обработки речевых сигналов

Первые два компонента инструментария были рассмотрены в первой части данной работы. Целью данной работы является рассмотрение третьего компонента программного инструментария, предназначенного для оценки эффективности систем обработки речевых сигналов по таким критериям как качество и разборчивость речи, достоверность автоматического распознавания речи.

1. Оценивание качества речи

Такие критерии как качество и разборчивость речи достаточно давно используются при акустической экспертизе каналов речевой коммуникации. Качество речи является важным параметром для коммерческих линий связи, тогда как разборчивость речи используется преимущественно в военных каналах речевой коммуникации.

Качество и разборчивость речи связаны между собой, однако эта связь не является однозначной: повышение качества речи приводит к росту разборчивости речи, однако обратное утверждение, вообще говоря, неверно. Качество речи удобно оценивать в соответствии со стандартом ITU-T PESQ/P.862, действовавшим до

2010 года. Причин тому несколько. Во-первых, существуют программы моделирования алгоритма PESQ в среде Matlab, а во-вторых, далеко не все провайдеры успели перейти на новый стандарт ITU-T POLQA/P.863 (разработанный в период 2006-2011 гг.).

Модель алгоритма PESQ, разработанная для среды Matlab, представлена в работе [3] в виде р-кодов.

Синтаксис функции `pesq()`, реализующей алгоритм PESQ:

```
[pesq_mos]=pesq(sfreq,cleanfile.wav,enhanced.wav)
```

где 'sfreq' – частота дискретизации в Гц (8000 или 16000 Гц);

'cleanfile.wav' – файл образцового речевого сигнала (без шума);

'enhanced.wav' – файл речевого сигнала на выходе системы шумоподавления.

На рис. 2 показаны результаты измерений параметра PESQ для чистого речевого сигнала (а), зашумленного сигнала (б), а также сигналов, восстановленных с помощью винеровской (в) и \log MMSE фильтрации (г). Данные результаты свидетельствуют о работоспособности предложенной в [3] модели алгоритма PESQ.

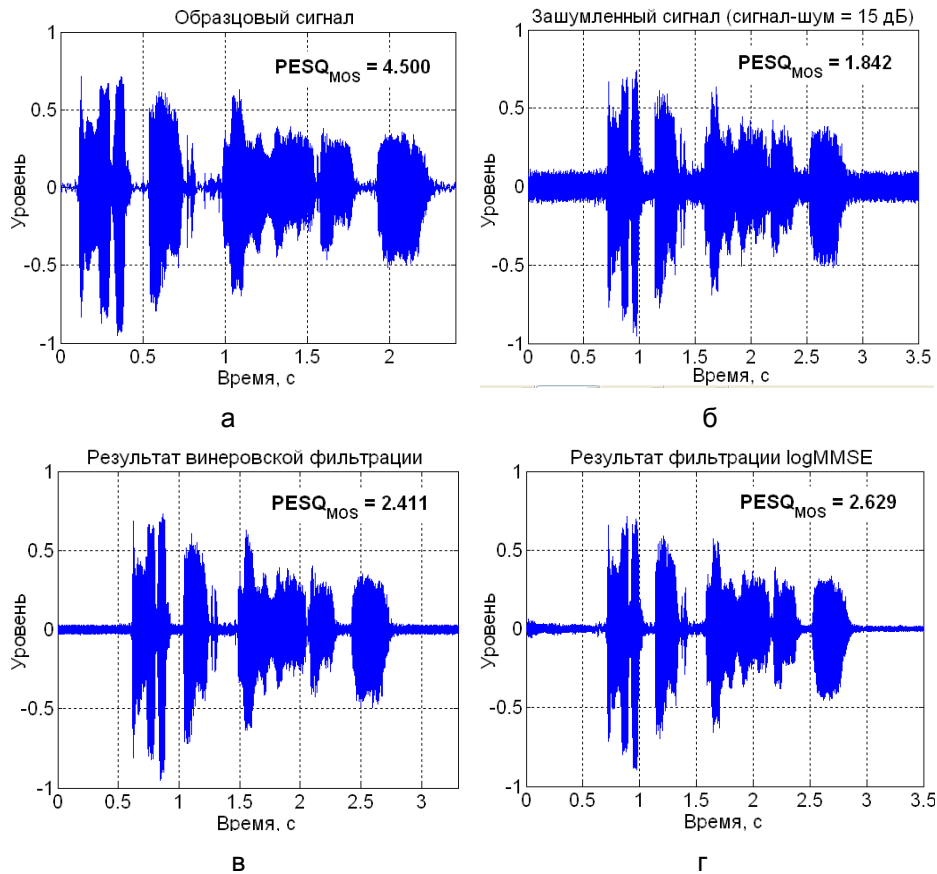


Рис. 2. Измерение $PESQ_{MOS}$ для чистого речевого сигнала (а), зашумленного сигнала (б) и сигнала, восстановленного путем винеровской (в) и logMMSE фильтрации (г)

2. Оценивание разборчивости речи

Если кроме шумовой помехи имеется реверберация, тогда для оценивания разборчивости речи необходимо использовать формантно-модуляционный (ФМ) метод, согласно которому [1, 2], как и в формантном методе [3], вычисляют артикуляционную разборчивость A , представляющую собой 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 в том, что в форматном 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:

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

где SNR_k – парциальное отношение сигнал-шум; D_{sk} и D_{nk} – соответственно дисперсии стационарных сигнала и шума в k -й полосе частот. В ФМ методе для оценивания E_k используют способ, применяющийся в модуляционном методе и позволяющий корректно учесть влияние реверберационной помехи [4, 5]. 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

$y(t) = \int_{-\infty}^{\infty} h(v)x(t-v)dv + n(t)$, где $n(t)$ – шумовая помеха, $h(v)$ – импульсная характеристика (ИХ) канала передачи, 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:

$$E_k = \frac{1}{14} \sum_{i=1}^{14} SNR_{ik}, \quad (5)$$

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

где $m_k(F_i)$ – коэффициент модуляции дисперсии $D_y(t)$ сигнала $y(t)$, наблюдаемого в k -й полосе частот [4, 5]. Коэффициент модуляции $m_k(F_i)$ оценивают с использованием преобразования Фурье [1]:

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

где $||$ – 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 [8] предложена ускоренная версия ФМ метода, в которой вместо 14 измерительных опытов с тестовыми сигналами (3) достаточно поставить единственный опыт с тестовым сигналом вида:

$$x(t) = \xi(t)\sqrt{f_5(t)},$$

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

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

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:

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

При компьютерном моделировании измерительной системы, реализующей ФМ метод, целесообразно создать унифицированную систему вычислительных алгоритмов и программных модулей, позволяющую реализовывать оценивание разборчивости речи как по алгоритмам ФМ метода, так и по алгоритмам формантного метода. Приведенные в работе [1] результаты аналитических и экспериментальных исследований форматного и формантно-модуляционного методов измерений свидетельствуют, что в случае преобладания шумовой помехи формантный метод предпочтителен, поскольку является наиболее точным и быстрым. Полная версия ФМ метода проигрывает ему по точности измерений разборчивости речи в 1,5...2 раза, а по времени измерений – в 14 раз. Ускоренная версия ФМ метода требует такого же времени измерений, что и формантный метод, проигрывая по точности в 6...8 раз.

3. Оценивание достоверности автоматического распознавания речи

Предлагаемый в данной работе инструментальный комплекс Matlab-НТК состоит из двух частей - в дальнейшем будем именовать эти части комплекса «Matlab» и «НТК», поскольку эти части содержат файлы-программы вычислительных сред Matlab и НТК, соответственно.

Часть «Matlab» призвана облегчить проведение вычислительных экспериментов, автоматизируя процесс вычислений как в среде Matlab, так и вне этой среды. При этом основным (управляющим) является специальный скрипт среды Matlab.

Часть «НТК» комплекса является самостоятельной в том смысле, что может эксплуатироваться автономно. При этом изменениям обычно подвергаются наборы тестовых и обучающих wav-файлов, словарь транскрипций и конфигурационный файл.

Часть «НТК» комплекса

Данная часть комплекса состоит из раздела bin, в котором находятся исполняемые файлы - компоненты пакета НТК, а также подразделы

bin\php (файлы процессора PHP-скриптов) и bin\scripts (рабочие скрипты php).

В разделе input\ находятся входные данные: файл конфигурации config; словарь транскрипций слов dict.trn; список слов, присутствующих в тестовых сигналах wordlist; директории, содержащие наборы тестовых и обучающих данных.

Файл конфигурации config идентичен стандартному файлу конфигурации НТК [4], за исключением двух новых параметров: количества вариантов произнесения одного слова, а также названия директории с набором тестовых и обучающих данных.

Файл wordlist содержит список всех слов, встречающихся в тестовых сигналах.

Файл dict.trn содержит список распознаваемых слов и их фонетические транскрипции.

В разделе output\ хранятся промежуточные данные, генерируемые скриптами и компонентами НТК.

В разделе result\ хранятся оценки качества распознавания речевых сигналов.

Работу с частью «НТК» комплекса можно представить в виде последовательности следующих этапов.

Этап 1. Выбрать произвольное имя для набора тестовых данных. Пусть, например, это будет «data_1».

Этап 2. Записать обучающие и тестовые данные и разместить их в соответствующие директории «input\data_1\train\» и «input\data_1\test\» в вышеописанном представлении.

Этап 3. Составить файл «input\dict.trn»

Этап 4. Составить файл «input\wordlist»

Этап 5. Отредактировать файл конфигурации «input\config», вставив новые параметры ZIMPLCOUNT (количество используемых при обучении вариантов произнесения каждого слова) и ZWAVNAME (название папки с набором речевых данных).

Этап 6. Последовательно запустить на выполнение пять .bat-файлов: 1) 0_clear.bat (производит «очистку» - удаление предыдущих результатов и предварительных данных); 2) 1_preprocess.bat (подготавливает данные - выполняет предварительную обработку и преобразование данных в НТК-совместимый формат); 3) 2_train.bat (производит обучение системы распознавания НТК); 4) 3_test.bat (производит непосредственное распознавание); 5) 4_results.bat (производит обработку результатов распознавания и выдачу результатов).

В результате распознавания речевого сигнала в файл «result\<имя набора дан-

ных>\out.txt» будут записаны оценки качества распознавания.

Часть «Matlab» комплекса

Данная часть комплекса состоит из основного управляющего m-скрипта test.m и четырех разделов: 1) processors\ (хранятся новые блоки обработки); test\ (хранятся управляющие m-скрипты); 2) resq\ (хранятся r-скрипты библиотеки вычисления параметра PESQ); 3) voicebox\ (хранятся m-скрипты библиотеки Voicebox); 4) jsonlab\ (скрипты для работы с форматом JSON).

Работу с частью «Matlab» комплекса можно представить в виде последовательности следующих этапов.

Этап 1. Копирование исходного набора данных в промежуточный набор данных.

Этап 2. Искажение образцов речевых сигналов в промежуточном наборе.

Этап 3. Копирование промежуточного набора в окончательный набор.

Этап 4. Коррекции речевых сигналов из окончательного набора.

Этап 5. Присвоение переменной ZWAVNAME (в файле конфигурации) значения, соответствующего имени окончательного набора.

Этап 6. Выполнение процедуры распознавания искаженной речи путём последовательного вызова и исполнения необходимых .bat-файлов (см. выше).

Этап 7. Чтение и анализ содержимого файла «result.json» из папки с результатами.

Особенности работы с комплексом Matlab-НТК

Для эксплуатации комплекса Matlab-НТК характерен ряд особенностей. Во-первых, активно используется такой тип переменных, как структура.

Пример работы со структурой:

```
% создание структуры и задание изначальных значений
test = struct('operand1', 2, 'operand2', 3);
test.operand1 + test.operand2 % отображение значения "5"
```

Во-вторых, в качестве блоков обработки сигналов используются структуры с определёнными названиями переменных.

Пример задания блока обработки:

```
% создание структуры обработчика
block = struct();
% задание zrandn в качестве функции обработки
block.method = @zrandn;
```

% задание свойств обработчика

```
block.parameters = struct('std', 0.02, 'mean', 0);
```

В строке `block.method = @zrandn` символ `@` означает передачу функции по ссылке вместо её непосредственного вызова. Это позволяет сохранить функцию в переменную `method` структуры. Таким образом, структура блока обработки должна обязательно включать в себя две переменные: `method` (ссылка на функцию обработки); `parameters` (структура с описанием параметров фильтра).

В целом, такой способ задания блоков обработки позволяет добиться единого интерфейса работы с каждым блоком, что позволит легко добавлять новые блоки, не заботясь о способе вызова каждой конкретной функции обработки. Такой же способ передачи параметров используется и в библиотеке Voicebox [6].

Так как части «НТК» и «Matlab» находятся в разных директориях, для удобства их совместной эксплуатации создана функция `zsystem()`, представляющая собой «обертку» для функции `system()`, отличающуюся тем, что она выполняет команды над файлами директории НТК (содержащей исполняемые файлы инструментария НТК, звуковые файлы, файлы с текстовыми транскрипциями и т.п.).

Кроме того, для эксплуатации комплекса Matlab-НТК разработан ряд других m-функций, среди которых ведущую роль играют функции `runtest()` и `printresults()`. Функция `runtest()` является основной функцией комплекса – ее входными данными является переменная `params`, определяющая структуру конфигурации, а выходными данными является структура `results`, содержащая результаты распознавания и оценивания параметра PESQ.

Выводы

Предложенный программный инструментарий позволяет экспериментально оценивать

эффективность алгоритмов коррекции речевых сигналов, используя такие критерии как качество и разборчивость речи, вероятность правильного распознавания речи. Разработанный инструментарий является неоднородным: часть вычислений производится в среде Matlab, а часть – вне этой среды. Несмотря на указанную особенность, данный инструментарий позволяет совместно использовать все необходимые программные компоненты, что обеспечивает сквозную автоматизацию обработки речевых сигналов, начиная от их намеренного искажения, последующей коррекции – и заканчивая оценкой качества восстановленных речевых сигналов. Разработанный комплекс инструментарий может быть использован как для учебных целей, так и для научных исследований.

Литература

1. POLQA Application Guide. – Germany: HEAD acoustics GmbH, 2012.–29 p.
2. Recommendation ITU-T P.863. – Switzerland: Geneva, 2011. – 76 p.
3. Loizou P. Speech enhancement: Theory and Practice. – Boca Raton: CRC Press, 2007. – 632 p.
4. Simulation of Acoustic Environments Including the Transmission over Telephone Channels. [Электронный ресурс]. – Режим доступа к ресурсу: <http://dnt.kr.hsnr.de/sireac.html> (21.05.2013).
5. Springer Handbook of Speech Processing / Под редакцией Jacob Benesty, M. Mohan Sondhi, Yiteng Huang. – Berlin: Springer-Verlag, 2008. – P. 1176.
6. VoiceBox: Speech Processing Toolbox for MATLAB. [Электронный ресурс]. – Режим доступа к ресурсу: <http://www.ee.ic.ac.uk/hp/staff/dmb/> (21.05.2013).

УДК 004.934

Програмний інструментарій для тестування систем обробки мовленнєвих сигналів. Частина 2. Оцінка ефективності

А.М. Продеус, д-р техн. наук

Національний технічний університет України «Київський політехнічний інститут»,
пр. Перемоги, 37, Київ-56, 03056, Україна.

Запропоновано склад програмного інструментарію для досліджень і оптимізації алгоритмів систем корекції й кодування мовленнєвих сигналів, що піддані різного роду спотворен-

ням, таким як шумова й ревербераційна перешкоди, помилки кодування. У першій частині даної роботи розглянуто інструментарій, що дозволяє розв'язувати такі завдання як створення корпусів зашумленої мови й корекція зашумлених мовленнєвих сигналів. У даній, другій частині роботи, запропоновано програмний інструментарій для атестації систем обробки мовних сигналів, що піддаються тестуванню, за такими критеріями як якість і розбірливість мовлення, достовірність автоматичного розпізнавання мовлення. Розроблений інструментарій є неоднорідним: частина обчислень виконується в середовищі Matlab, а частина - поза цим середовищем. Незважаючи на зазначену особливість, даний інструментарій дозволяє спільно використовувати всі необхідні програмні компоненти, що забезпечує наскрізну автоматизацію обробки мовленнєвих сигналів, починаючи від їхнього навмисного спотворення, наступної корекції - і закінчуючи оцінкою якості відновлених мовленнєвих сигналів. Бібл. 6, рис. 2.

Ключові слова: програмний інструментарій, критерії якості, тестування систем.

UDC 004.934

Software toolkit for testing of speech signals processing systems. Part 2. Efficiency estimation

A.N. Prodeus, Dr.Sc.

National Technical University of Ukraine «Kyiv Polytechnic Institute»,
37 Prospect Peremogy, Kiev 03056, Ukraine.

Structure of the software toolkit for the study and optimization of algorithms of correction and coding systems, which are subject of various distortions, such as noise and reverberation disturbance, coding errors is proposed. The first part of this paper was devoted to consideration of toolkit which permit solve such problems as creating of noisy speech corpora and correction of noisy speech signals. Software toolkit for the validation of test systems for processing speech signals using criteria such as the quality and intelligibility of speech, the accuracy of automatic speech recognition, was proposed in the second part of the work. A developed software toolkit is not uniform: some calculations done in an environment Matlab, and some - outside of this environment. Despite the specified feature, this toolkit allows user to combine all necessary software components, which provides through automation of the processing of speech signals from their deliberate distortions, subsequent correction - to evaluation of quality of the reconstructed speech signal. Reference 6, figures 2.

Keywords: software toolkit, quality criteria, systems testing.

References

1. (2012), "POLQA Application Guide". Germany. HEAD acoustics GmbH, P. 29.
2. (2007), "Recommendation ITU-T P.863". Switzerland. Geneva, P. 76.
3. Loizou P. (2007), "Speech enhancement: Theory and Practice". Boca Raton: CRC Press, p. 632.
4. Simulation of Acoustic Environments Including the Transmission over Telephone Channels. [On-line]. – Available: <http://dnt.kr.hsr.de/sireac.html> (21.05.2013).
5. Jacob Benesty, M. Mohan Sondhi, Yiteng Huang. (2008), "Springer Handbook of Speech Processing". Berlin: Springer-Verlag, P. 1176.
6. VoiceBox: Speech Processing Toolbox for MATLAB. [On-line]. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/> (21.05.2013)

Поступила в редакцію 27 мая 2013 г.