

Теория сигналов и систем

УДК 004.934

К.С. Замша, Б.В. Лозинский, Ю.А. Митяй, Е.С. Степановская, А.Н. Продеус, д. - р. техн. наук
Национальный технический университет Украины «Киевский политехнический институт»,
ул. Политехническая, 16, корпус 12, г. Киев, 03056, Украина.

Объективное и субъективное оценивание качества речевых сигналов с ограниченной полосой частот

Путем вычисления коэффициентов корреляции и построения карт соответствия, сопоставлены результаты объективного и субъективного оценивания качества речевых сигналов, искаженных из-за ограничения полосы частот. Показано, что карты соответствия, дополненные информацией о способе их построения, более полезны в прикладном плане, нежели коэффициенты корреляции, поскольку, во-первых, выполняют функцию калибровочного коэффициента, а во-вторых, допускают возможность нелинейной зависимости между объективными и субъективными оценками качества речевого сигнала. Библ. 7, рис. 5, табл. 3.

Ключевые слова: качество речевого сигнала; объективная оценка; субъективная оценка; карта соответствия; коэффициент корреляции.

Введение

Объективные (инструментальные) меры качества речевых сигналов предлагались различными исследователями, начиная с середины 70-х годов прошлого столетия. Наиболее полно вопрос оценки надежности этих мер освещен в [7]. При этом предложена следующая схема оценивания надежности объективных мер. Речевые сигналы высококачественной речевой базы подвергаются контролируемому искажению, после чего подвергаются объективному и субъективному тестированию. На завершающем этапе результаты объективного и субъективного тестирования сопоставляются путем оценивания коэффициента корреляции.

Анализируя работы, касающиеся темы объективного оценивания качества речи, приходится с удивлением констатировать, что многие исследователи [1,3,4] безоговорочно следуют этим рекомендациям, не обращая внимания на то обстоятельство, что коэффициент корреляции недостаточно информативен в случае не-

линейной зависимости результатов субъективной и объективной акустической экспертизы.

Между тем, значительно более полезным в прикладном плане, на наш взгляд, является построение карты соответствия результатов субъективного и объективного тестирования [5]. Такую карту строят в несколько этапов. Сначала строят зависимости значений субъективного и объективного показателей от некоторого параметра θ , определяющего степень искажений речевого сигнала. Например, в случае шумовой аддитивной помехи в качестве параметра θ может служить отношение сигнал-шум или характер окрашенности шума, для реверберации это время реверберации, для цифровых линий связи - уровень битовой ошибки, вид используемого кодека и т.д. Затем, используя параметр θ как связующее звено, строят карту соответствия результатов субъективных и объективных измерений. Фактически это график, позволяющий пересчитывать результаты объективных измерений в результаты субъективного оценивания. В [5] показано, что такая карта соответствия может иметь вид нелинейной функциональной зависимости.

Цель данной работы состоит в обосновании предпочтительности использования карт соответствия, вместо коэффициентов корреляции, при сопоставлении результатов объективного и субъективного оценивания качества речевого сигнала. Другая цель работы состоит в выработке предложений по развитию методики построения таких карт соответствия.

Для определенности, в данной работе рассмотрена ситуация, когда искажающим фактором является ограниченность полосы частот канала речевой коммуникации. Однако полученные при анализе этой ситуации результаты носят достаточно общий характер, и поэтому могут быть использованы едва ли не при любых видах искажений речевых сигналов.

1. Субъективное и объективное оценивание качества сигналов

В данной работе использован метод субъективного тестирования, именуемый «оценкой степени деградации» (Degradation Category Rating - DCR). Согласно этому методу, слушатели на слух сравнивают качество искаженного и эталонного речевых сигналов, выставляя оценки по пятибалльной шкале, именуемой «шкалой средней деградации» (Degradation Mean Opinion Score - DMOS) [3].

В качестве объективных мер качества речевого сигнала использованы пять показателей, таких как сегментное отношение сигнал-шум SSNR (Segmental Signal to Noise Ratio), логарифмическое спектральное искажение LSD (Logarithmic Spectral Distortion), частотно-взвешенное сегментное отношение сигнал-шум fwSSNR (frequency-weighted segmental SNR), барк-спектральное искажение BSD (Bark Spectral Distortion) и перцептуальная оценка качества речи PESQ (Perceptual Evaluation of Speech Quality) [3].

Мера SSNR относится к группе показателей временной области:

$$SSNR = \frac{1}{M} \sum_{m=1}^M 10 \lg \left[\frac{\sum_{n=R(m-1)+1}^{Rm} x^2(n, m)}{\sum_{n=R(m-1)+1}^{Rm} [x(n, m) - y(n, m)]^2} \right], (1)$$

где $x(n, m)$ и $y(n, m)$ - n -е выборки m -го фрейма чистого сигнала $x(n)$ и искаженного сигнала $y(n)$, соответственно; M - количество фреймов; R - количество выборок в фрейме.

Мера LSD принадлежит к группе показателей частотной области:

$$LSD = \frac{2}{RM} \sum_{m=1}^M \sum_{j=1}^J |G\{X(j, m)\} - G\{Y(j, m)\}|, (2)$$

$$G\{X(j, m)\} = \max\{20 \lg(|X(j, m)|), \delta\},$$

$$\delta = \max_{l, k} \{20 \lg(|X(j, m)|)\} - 50,$$

где $X(j, m)$ и $Y(j, m)$ - дискретные преобразования Фурье m -го фрейма сигналов $x(n)$ и $y(n)$, соответственно, j - номер частотной выборки, J - количество частотных выборок.

К группе показателей частотной области также принадлежит fwSSNR:

$$fwSSNR = \frac{10}{M} \sum_{m=1}^M \frac{\sum_{k=1}^K W(k, m) \lg \frac{|X(k, m)|^2}{(|X(k, m)| - |Y(k, m)|)^2}}{\sum_{k=1}^K W(k, m)}, (3)$$

где k - номер критической полосы; K - количество критических полос; $|X(k, m)|$ - амплитудный спектр чистого сигнала; $W(j, m)$ - весовые коэффициенты [4].

Мера BSD относится к группе перцептуальных показателей, учитывающих особенности слуховой системы человека:

$$BSD = \frac{\sum_{m=1}^M \sum_{k=1}^K [B_x(k, m) - B_y(k, m)]^2}{\sum_{m=1}^M \sum_{k=1}^K [B_x(k, m)]^2}, (4)$$

где $B_x(k, m)$ и $B_y(k, m)$ - барк-спектры m -го фрейма сигналов $x(n)$ и $y(n)$, соответственно, k - номер критической полосы частот.

Описание меры PESQ, также относящейся к перцептуальным показателям, весьма громоздко, в сокращенном виде его можно найти в [3,4]. В данной работе использована «широкополосная» версия WB-PESQ, позволяющая оценивать качество речевых сигналов в полосе частот от 50 Гц до 7 кГц.

2. Организация экспериментальных исследований

Запись эталонных речевых сигналов для 4-х дикторов-женщин и 4-х дикторов-мужчин, читающих текст по юридической тематике, производилась в заглушенном помещении кафедры акустики НТУУ «КПИ», при частоте дискретизации 22050 Гц и битовой глубине 16 бит. Время реверберации в помещении составляло 0,15 с, отношение сигнал-шум было близким 38 дБ.

Искажение речевого сигнала производилось путем поочередного пропуска его через нерекурсивные фильтры низкой частоты, синтезированные методом Ремеза. Характеристики фильтров:

- частота среза от 0,5 до 10,5 кГц с шагом 0,5кГц;
- размер переходной зоны 5% от полосы пропускания;
- неравномерность АЧХ в полосе пропускания 1 дБ;
- затухание в полосе подавления минус 80дБ.

Субъективное оценивание качества речевого сигнала производилось 17 аудиторами, юношами и девушками в возрасте 21-22 лет, с использованием фрагментов сигналов протяженностью не более 15 секунд. Объективное оценивание показателей SSNR, LSD, fwSSNR, BSD и PESQ осуществлялось в среде Matlab и производилась по фрагментам сигналов протяженностью 1 минута.

3. Результаты экспериментов

Усредненные результаты субъективного оценивания, приведенные на рис. 1, свидетельствуют о практически монотонном возрастании качества сигнала с увеличением полосы частот.

Результаты объективного оценивания качества этих же речевых сигналов приведены в [2,6]. Фрагменты этих результатов приведены на рис. 2.

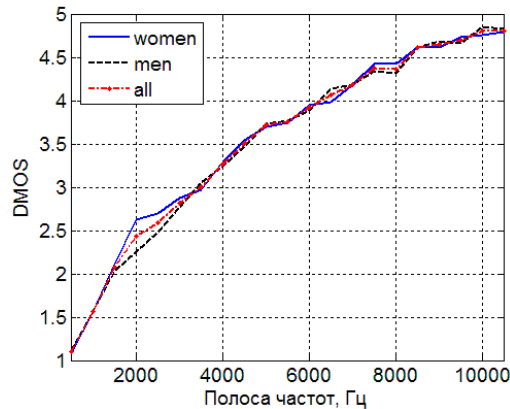
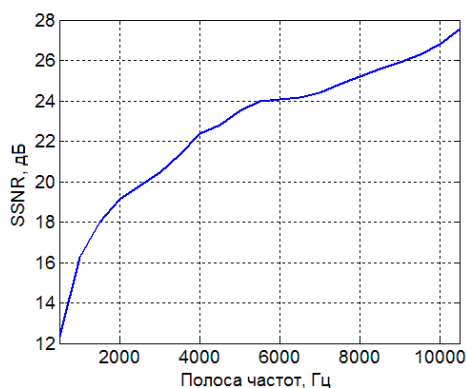


Рис. 1. Изображение результатов субъективного оценивания по шкале DMOS

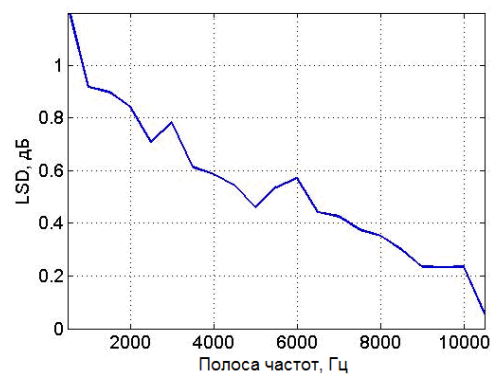
Значения коэффициентов корреляции между результатами объективного и субъективного оценивания приведены в табл. 1, а карты соответствия объективных и субъективных оценок представлены на рис. 3.

В отличие от [5], где карты соответствия строились простым соединением прямыми линиями точек на плоскости, в данной работе построение карт соответствия, приведенных на

рис. 3, произведено двумя способами: с помощью линейной и полиномиальной регрессии. В табл. 2 приведены значения максимальных ошибок аппроксимации. Очевидно, разумным следует считать выбор аппроксимирующего полинома минимального порядка p , при котором достигается приемлемая для приложений ошибка аппроксимации.



(а)



(б)

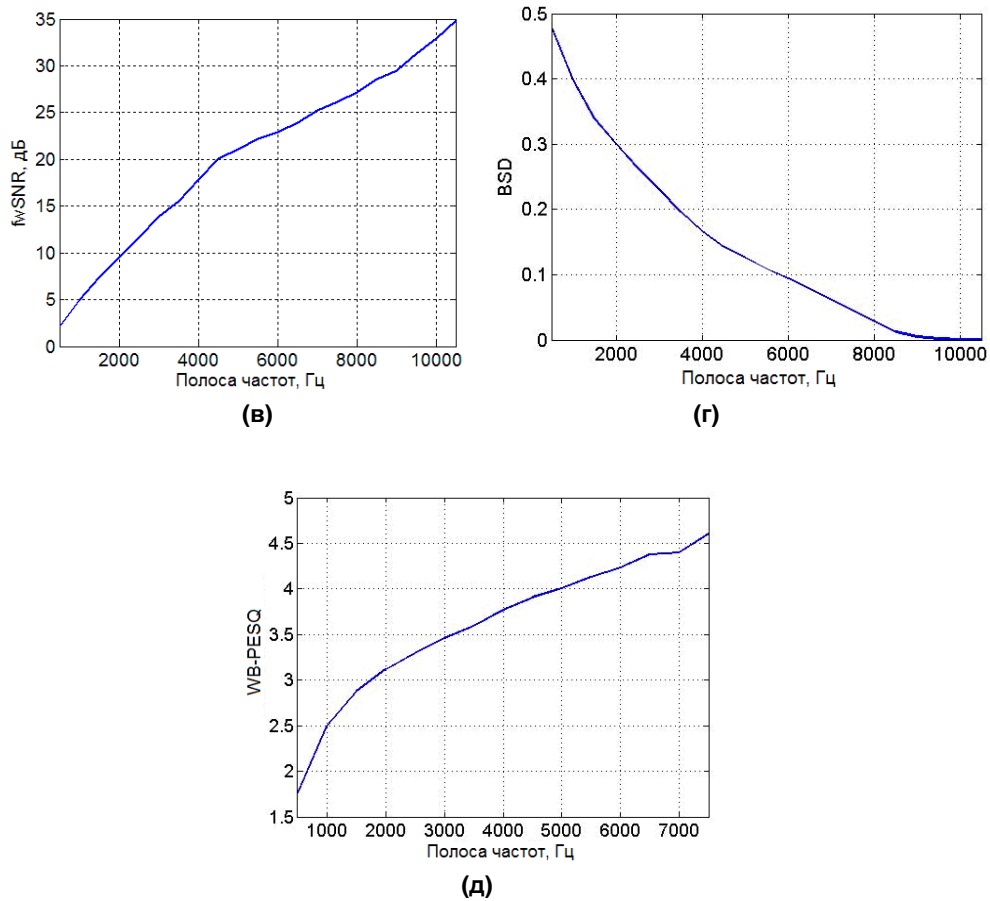


Рис. 2. График зависимости значений объективных показателей качества речи от полосы частот

Таблица 1. Коэффициенты корреляции объективных и субъективных оценок

Объективный показатель		SSNR	LSD	fwSSNR	BSD	WB-PESQ
Коэффициент корреляции	Жен	0,979	-0,959	0,979	-0,994	0,963
	Муж	0,968	-0,935	0,993	-0,981	0,941
	Все	0,979	-0,970	0,990	-0,997	0,962

Как следует из табл. 1, 2 и рис. 3, наибольшие значения коэффициента корреляции получены для показателей fwSSNR и BSD. При этом, как следует из табл. 2, максимальная ошибка

аппроксимации не превышает 0,1 для зависимости DMOS(fwSSNR) при порядке аппроксимирующего полинома $p = 4$ и не превышает 0,112 для зависимости DMOS(BSD) при $p = 3$.

Таблица 2. Максимальная ошибка аппроксимации

Объективный показатель Порядок полинома p	SSNR	LSD	fwSSNR	BSD	WB-PESQ
1	0,469	0,634	0,407	0,144	0,542
2	0,382	0,538	0,156	0,130	0,544
3	0,247	0,388	0,180	0,112	0,493
4	0,245	0,389	0,100	0,112	0,494
5	0,243	0,386	0,098	0,129	0,440

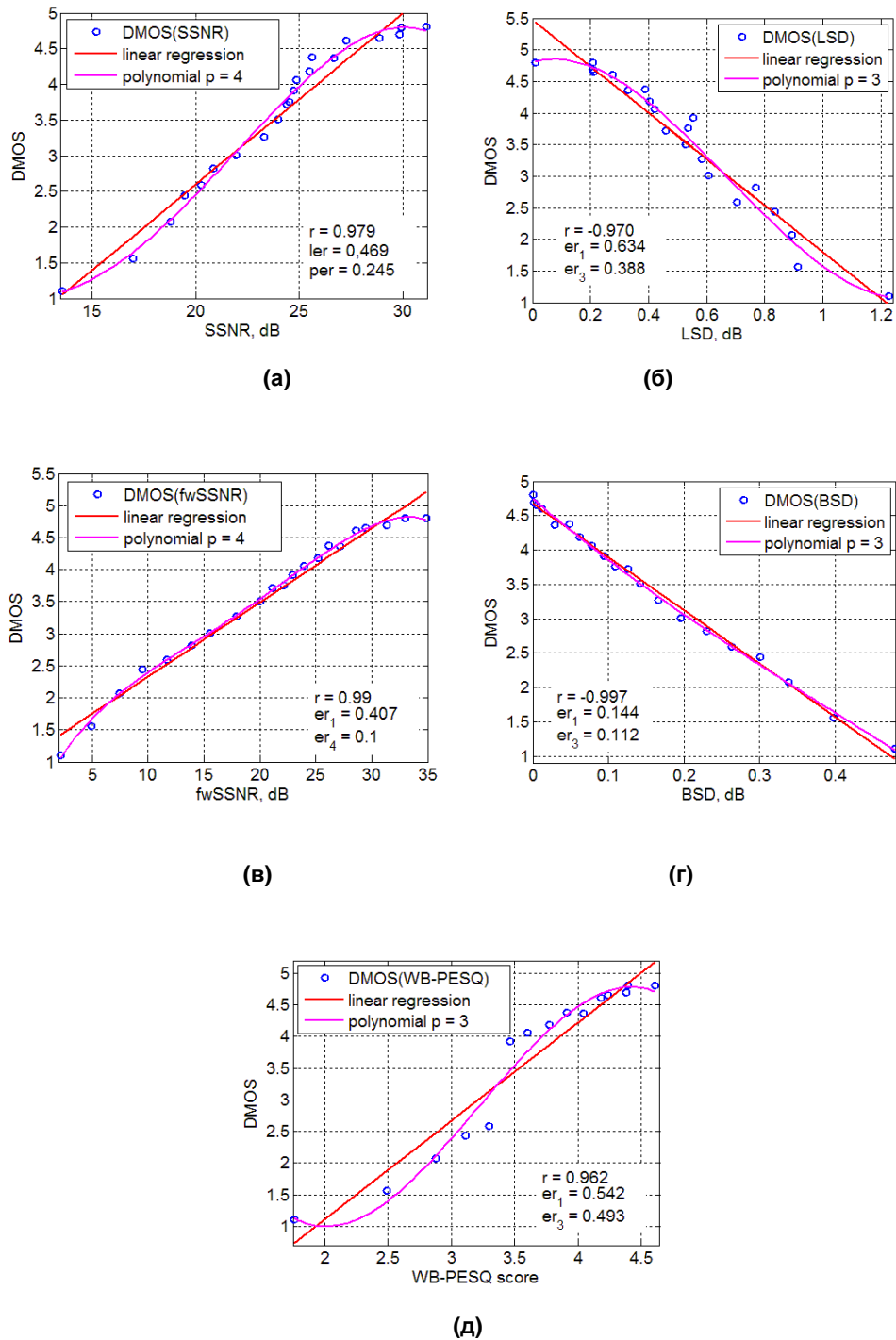


Рис. 3. Аппроксимация зависимостей между результатами объективного и субъективного оценивания качества речевых сигналов

Весьма неожиданными оказались результаты для показателя WB-PESQ, являющегося стандартом де-факто при оценивании качества речевых сигналов в линиях связи. На рис. 3д хорошо видно, что, несмотря на высокое значение коэффициента корреляции ($r = 0,962$), зависимость между результатами объективных и субъективных испытаний носит выраженный

нелинейный характер, вследствие чего максимальная ошибка аппроксимации составила $er_1 = 0.542$ при линейной аппроксимации и $er_3 = 0.493$ при аппроксимации полиномом 3-го порядка.

Обнаружилось также, что использование

полиномиальной аппроксимации не позволяет адекватно описать выявленную нелинейную зависимость. Как следует из рис. 3д и 4а, повышение порядка аппроксимирующего полинома с $p=3$ до $p=6$ не позволяет уменьшить ошибку аппроксимации. Более того, такое повышение негативно сказывается на форме аппроксими-

рующей функции, нарушая ее явно монотонный характер. И лишь использование кубических сплайнов позволило повысить качество аппроксимации (рис. 4б), снизив максимальную ошибку аппроксимации до 0,1 и одновременно позволяя сохранить монотонность аппроксимирующей функции.

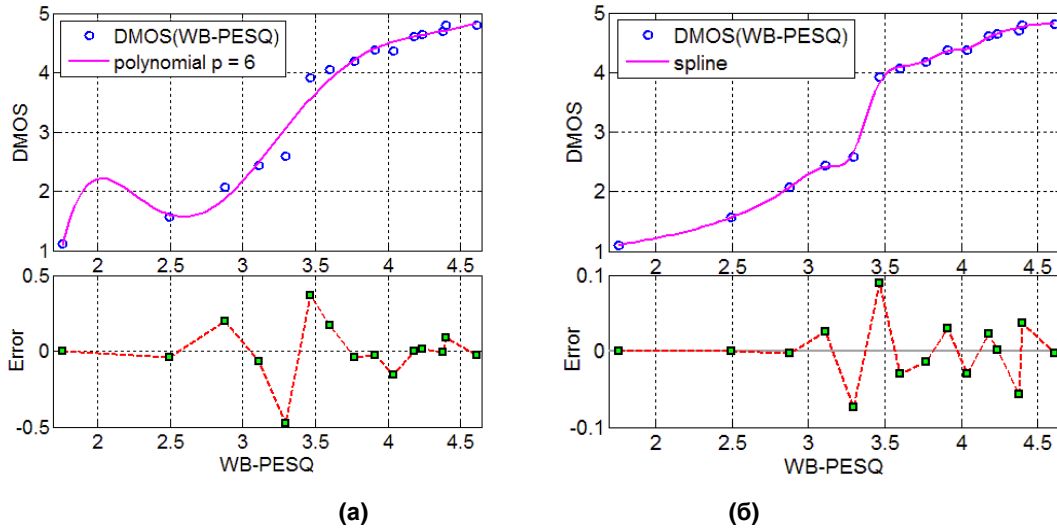


Рис. 4. Аппроксимация зависимости DMOS(WB-PESQ): полиномом 6-го порядка (а) и кубическими сплайнами (б)

Результаты аппроксимации, представленные на рис. 3 и 4, можно использовать в графической форме. Для практического использования, однако, удобнее представить эти результаты в аналитической форме. Как следует из табл. 3, при полиномиальной аппроксимации достаточно ограничиться значением $p=4$.

Значения коэффициентов a_0, \dots, a_4 аппроксимирующего полинома

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$$

приведены в табл. 3.

Таблица 3. Коэффициенты аппроксимирующих полиномов

Объективный показатель Коэффициент полинома	SSNR	LSD	fwSSNR	BSD
a_0	3,9135	4,7967	0,4378	4,7478
a_1	-0,4484	1,3495	0,3291	-9,6804
a_2	0,0081	-9,1482	-0,0202	8,7368
a_3	0,0011	4,6923	0,0008	-14,0280
a_4	-3e-05	-0,1137	-1e-05	9,8195

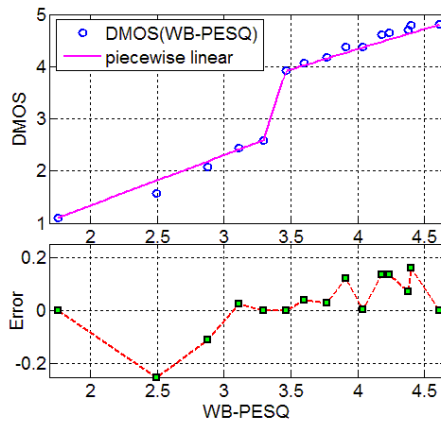
Поскольку использование полиномов не позволяет качественно аппроксимировать зависимость DMOS(WB-PESQ), необходимо альтернативное решение. Таковым может быть, например, кусочно-линейная аппроксимация, одним из вариантов которой является соотношение

$$y = \begin{cases} 0,9635 \cdot x - 0,5886, & 1,75 \leq x < 3,3; \\ 7,9456 \cdot x - 23,6, & 3,3 \leq x < 3,46; \\ 0,7732 \cdot x + 1,24, & 3,46 \leq x \leq 4,6, \end{cases} \quad (1)$$

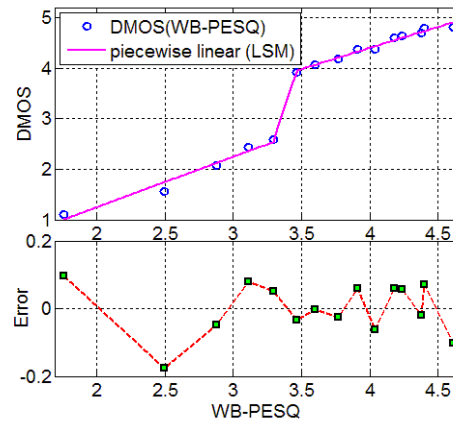
согласно которому четыре точки с координатами (1,76; 1,10), (3,30; 2,59), (3,46; 3,92) и (4,61; 4,80) соединяются прямыми линиями (рис. 5 а). Как следует из рис. 5а, при таком подходе удастся снизить ошибку аппроксимации до 0,25.

Еще больше, до 0,18, можно понизить ошибку аппроксимации, заменив первую и третью строки выражения (1) уравнениями линейной регрессии (рис. 5 б):

$$y = \begin{cases} 0,9934 \cdot x - 0,7369 & 1,75 \leq x < 3,3; \\ 7,9456 \cdot x - 23,6, & 3,3 \leq x < 3,46; \\ 0,8328 \cdot x + 1,0673 & 3,46 \leq x \leq 4,6, \end{cases} \quad (2)$$



(a)



(б)

Рис. 5. Кусочно-линейная аппроксимация зависимости DMOS(WB-PESQ): упрощенная (а) и с использованием метода наименьших квадратов (б)

Несомненным преимуществом карты соответствия перед коэффициентом корреляции является то, что такая карта не только описывает характер зависимости между субъективными и объективными оценками, но и служит калибровочным коэффициентом для пересчета результатов объективного оценивания качества речевого сигнала в результаты субъективного оценивания.

Поскольку результаты построения карты соответствия существенно зависят от выбранного способа построения, полученную карту соответствия весьма желательно дополнять информацией о способе ее построения. Так, например, если построение карты соответствия осуществлено с использованием методов аппроксимации, целесообразно к полученной карте прилагать оценку ее надежности в виде ошибки аппроксимации.

Выводы

Коэффициент корреляции, как мера надежности объективных показателей качества речевых сигналов, недостаточно информативен в случае возможной нелинейной зависимости между результатами субъективной и объективной акустической экспертизы. Значительно бо-

Таким образом, проблемы, возникшие при анализе соответствия показателей WB-PESQ и DMOS, свидетельствуют, в первую очередь, о низкой информативности коэффициента корреляции, обусловленной ошибочным предположением о линейной зависимости между результатами объективного и субъективного оценивания качества речевых сигналов.

более корректной и, вместе с тем, значительно более полезной для инженерных приложений альтернативой является карта соответствия результатов субъективной и объективной акустической экспертизы, дополняемая информацией о способе формирования такой карты.

Хотя в данной работе рассмотрена конкретная задача оценки качества речевого сигнала, ограниченного по полосе частот, полученные результаты носят достаточно универсальный характер и могут быть распространены не только на иные виды искажений речевых сигналов, но и на задачу оценивания разборчивости речи.

Список использованных источников

1. *Barnwell T.* Objective measures for speech quality testing. *J. Acoust. Soc. Am.* 66(6), Dec. 1979 - P. 1658-1663.
2. *Bogdanova N., Prodeus A.* Objective quality evaluation of speech band-limited signals. *Electronics and Communications*, Vol.19, No. 6(83), 2014. - P.58-65.
3. *Cote N.* Integral and diagnostic intrusive prediction of speech. - Springer-Verlag Berlin Heidelberg, 2011. – 267 p.
4. *Hu Y., Loizou P.* Evaluation of objective quality measures for speech enhancement. *IEEE*

- Transactions on Speech and Audio Processing, 16(1), 2008. - P. 229-238.
5. *Jiang W., Schulzrinne H.* Speech Recognition Performance as an Effective Perceived Quality Predictor. IEEE Int. Workshop on Quality of Service, pp. 269-275, 2002.
 6. *Prodeus A.* Reducing Sensitivity of Segmental Signal-to-Noise Ratio Estimator to Time-Alignment Error. International Journal of Electrical and Electronic Science, Vol. 2(2), August 20, 2015. - P. 31-36.
 7. *Quackenbush S., Barnwell T., Clements M.* Objective Measures of Speech Quality, Englewood Cliffs, NJ: Prentice-Hall, 1988.

Поступила в редакцию 24 ноября 2015 г.

УДК 004.934

К.С. Замша, Б.В. Лозинський, Ю.А. Митяй, Є.С. Степановська, А.М. Продеус, д .- р. техн. наук
Національний технічний університет України «Київський політехнічний інститут»,
вул. Політехнічна, 16, корпус 12, м. Київ, 03056, Україна.

Об'єктивне й суб'єктивне оцінювання якості мовленнєвих сигналів з обмеженою смугою частот

Шляхом обчислення коефіцієнтів кореляції й побудови карт відповідності, зіставлено результати об'єктивного й суб'єктивного оцінювання якості мовленнєвих сигналів, спотворених через обмеження смуги частот. Показано, що карти відповідності, доповнені інформацією про спосіб їх побудови, є більш корисними в прикладному плані, ніж коефіцієнти кореляції, оскільки, по-перше, виконують функцію калібрувального коефіцієнту, а по-друге, припускають можливість нелінійної залежності між об'єктивними й суб'єктивними оцінками якості мовленнєвого сигналу. Бібл. 7, рис. 5, табл. 3.

Ключові слова: *якість мовленнєвого сигналу; об'єктивна оцінка; суб'єктивна оцінка; карта відповідності; коефіцієнт кореляції.*

УДК 004.934

K. Zamsha, B. Lozynskiy, J. Mytiay, E. Stepanovskaya, A. Prodeus, Dr.Sc.
National Technical University of Ukraine "Kyiv Polytechnic Institute",
off. 233, Politekhnichna Str., 16, Kyiv, 03056, Ukraine

Objective and subjective assessment of bandlimited signaling speech quality

The coefficients of correlation and matching maps of the results of objective and subjective assessment of the speech quality for signals distorted due to the bandwidth limitations had been evaluated. It is shown that matching maps are much more useful in comparison with correlation coefficients because, firstly, they perform the function of the calibration factor and, secondly, they allow the possibility of non-linear dependence between objective and subjective speech quality assessments. Reference 7, figures 5, table 3.

Keywords: *speech signal quality; objective measure; subjective measure; matching map; correlation coefficient.*

References

1. *Barnwell, T.* (1979). Objective measures for speech quality testing. J. Acoust. Soc. Am. 66(6), Dec. Pp. 1658-1663.
2. *Bogdanova, N., Prodeus, A.* (2014). Objective quality evaluation of speech band-limited signals. Electronics and Communications, Vol.19, No. 6(83), Pp.58-65.
3. *Cote, N.* (2011). Integral and diagnostic intrusive prediction of speech. - Springer-Verlag Berlin Heidelberg, P. 267.
4. *Hu, Y., Loizou, P.* (2008). Evaluation of objective quality measures for speech enhancement. IEEE Transactions on Speech and Audio Processing, 16(1), Pp. 229-238.

5. *Jiang, W., Schulzrinne, H.* (2002). Speech Recognition Performance as an Effective Perceived Quality Predictor. IEEE Int. Workshop on Quality of Service, pp. 269-275.
6. *Prodeus, A.* (2015). Reducing Sensitivity of Segmental Signal-to-Noise Ratio Estimator to Time-Alignment Error. International Journal of Electrical and Electronic Science, Vol. 2(2), August 20, Pp. 31-36.
7. *Quackenbush, S., Barnwell, T., Clements, M.* (1988). Objective Measures of Speech Quality, Englewood Cliffs, NJ: Prentice-Hall.