Теория сигналов и систем

УДК 004.934

И.В. Котвицкий, **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 и музыкальных сигналов, подвергнутых фазовым искажениям

Недавние предварительные исследования показали, что для слуховой системы человека являются приемлемыми фазовые искажения музыкальных сигналов, если максимальная разница групповых времен задержки тракта в области высоких и низких частот не превышает 70 мс. Для речевых сигналов эта величина меньше и близка 50 мс. В данной работе получены более точные субъективные оценки зависимости качества речи и музыки от разницы времен групповой задержки. Кроме того, построены карты соответствия результатов объективного и субъективного оценивания качества искаженных сигналов. Показано, что при определенных условиях такие карты имеют выраженный нелинейный характер. Библ. 7, рис. 7, табл. 1.

Ключевые слова: *фазовые искажения; субъективная оценка; качество речевого сигнала; качество музыкального сигнала; показатели качества.*

Введение

В работе [1] рассмотрена модель возникновения фазовых искажений сигнала в гребенке цифровых нерекурсивных фильтров с сумматором (рис. 1). Гребенки цифровых фильтров широко используются в системах записи и воспроизведения, кодирования и декодирования, в линиях связи, в системах коррекции слуха [2, 3]. Между тем, если нерекурсивные фильтры, образующие гребенку фильтров, имеют разный порядок, фазовая частотная характеристика (ФЧХ) такой гребенки является нелинейной, что может существенно сказаться на качестве выходных сигналов. В работах [4, 5] показано, что для слуховой системы человека приемлемыми являются фазовые искажения речевых и музыкальных сигналов, если максимальная разница групповых времен задержки тракта в области высоких (8-11 кГц) и низких (90-180 Гц) частот не превышает 50-70 мс. Заметим, однако, что приведенные в [4, 5] субъективные оценки качества искаженных сигналов носили предварительный характер. Целью данной работы является уточнение этих оценок, а также построение карт соответствия между объективными и субъективными оценками качества речевых и музыкальных сигналов.

1. Модель возникновения фазовых искажений сигнала

Модель возникновения фазовых искажений сигнала, основанная на использовании гребенки октавных фильтров, предложена в [1] и модифицирована в [4]. При этом охватывалась типичная для речевых сигналов полоса частот 90-11000 Гц [4]. Основные параметры октавных фильтров, рассчитанных методом Ремеза, приведены в табл. 1, где f0 - центральная частота; Δf – полоса пропускания; n - порядок фильтра.

Таблица 1. Параметры гребенки октавных фильтров

$f0$, Гц	125	250	500	1000	2000	4000	8000
Δf , Гц	90	180	355	710	1400	2800	5600
n	4353	2903	2177	1320	927	545	437

Благодаря высокому порядку фильтров удалось добиться практически прямоугольной формы амплитудной частотной характеристики (АЧХ) каждого из фильтров гребенки, а в силу симметричности импульсных характеристик (ИХ) каждого из фильтров их ФЧХ линейные.

Однако суммарная ИХ такой гребенки октавных фильтров (рис. 1,а) несимметричная, а ФЧХ нелинейная (рис. 2,а), поскольку протяженности парциальных ИХ каждого из полосовых фильтров различны.

В качестве меры нелинейности ФЧХ $\Theta(f)$ линейной системы используем максимальную разницу $\Delta\tau_{max}$ значений группового времени задержки $\tau(f) = -d\Theta(f)/[2\pi df]$ (рис. 2,б), вычисленных для высоких (8-11 кГц) и низких (90-180 Гц) частот.

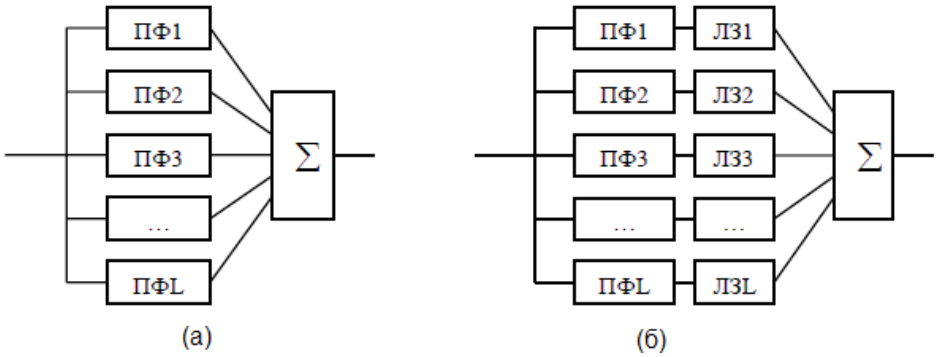
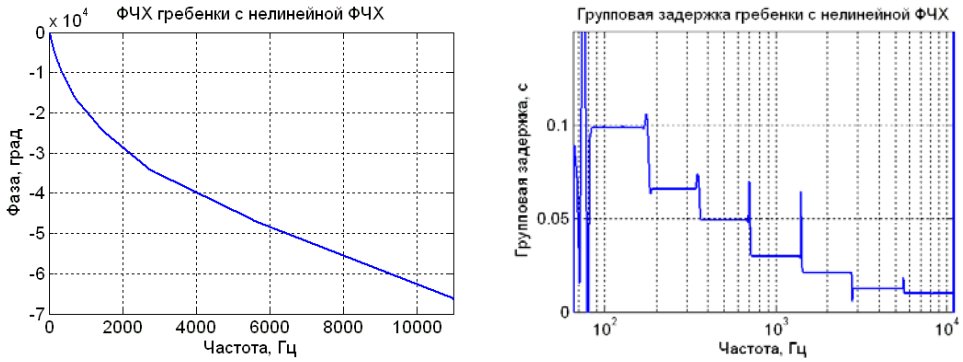


Рис. 1. Гребенка фильтров с сумматором (а) и с линиями задержки (б)



(а) **(б)**

Рис. 2. Графики ФЧХ (а) и $\tau(f)$ (б) для гребенки фильтров с сумматором

В данной работе, как и в [4, 5], рассмотрены две разновидности гребенок, а именно, с убывающей и возрастающей зависимостями $\tau(f)$. Конфигурацию ФЧХ, где ВЧ компоненты сигнала отстают от НЧ компонентов (рис. 2,б, «убывающее время задержки»), обозначим ФЧХ1. Конфигурация с «возрастающим временем задержки», обозначенная как ФЧХ2, получена добавлением линий задержки (ЛЗ) в каждый канал гребенки фильтров (рис. 1,б).

2. Оценивание качества акустического сигнала

При оценивании качества речевых сигналов использованы фрагменты, протяженностью 1 минута каждый, речевых сигналов для 4-х дикторов-женщин и 4-х дикторов-мужчин, читающих русский текст по юридической тематике. Запись сигналов произведена на кафедре акустики Национального технического университета «Киевский политехнический институт», в заглушенном помещении с временем реверберации 0,15 с, с частотой дискретизации 22050 Гц и битовой глубиной 16 бит.

При оценивании качества музыкальных сигналов использованы фрагменты восьми музыкальных произведений протяженностью 30-45 секунд каждый. При этом половина произведений принадлежала жанру классической музыки («Ave Maria» Дж. Каччини, «Этюд №4, соч. 10, Ф. Шопена», 5-я симфония П. Чайковского, увертюра «Фауст» Р. Вагнера), а половина – жанру популярной музыки («Mamamia» ABBA, «Shes_Leaving_Home» Beatles, «Я піду в далекі гори» К. Цісик, «Mademoiselle Hyde» L. Fabian).

Все сигналы записаны с частотой дискретизации 22050 Гц и битовой глубиной 16 бит.

В субъективном оценивании участвовало 32 человека, средний возраст которых составил 22 года. Оценивание качества сигналов производилось по шкале DMOS (Degradation Mean Opinion Score) [7], с использованием специально разработанной компьютерной программы, с помощью которой слушателю в случайном порядке предъявлялись искаженные сигналы для $\Delta\tau_{max} \approx 30$, 50, 70 и 90 мс, а также неискаженный сигнал ($\Delta\tau_{max} = 0$).

Для сопоставления результатов субъективного и объективного оценивания использовались результаты работ [4,5], где представлены оценки четырех мер качества: сегментного отношения сигнал-шум (Segmental Signal to Noise Ratio – SSNR), логарифмически-спектрального искажения (Logarithmic Spectral Distortion - LSD), барк-спектрального искажения (Bark Spectral Distortion – BSD) и перцептуального качества речи (Perceptual Evaluation of Speech Quality - PESQ) [6].

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

$$LSD = \frac{2}{KL}\sum_{l}\sum_{k=0}^{\frac{K}{2}-1} |G\{X(l,k)\} - G\{Y(l,k)\}|, \quad (2)$$

$$G\{X(l,k)\} = \max\{20\lg(|X(l,k)|), \delta\},$$

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

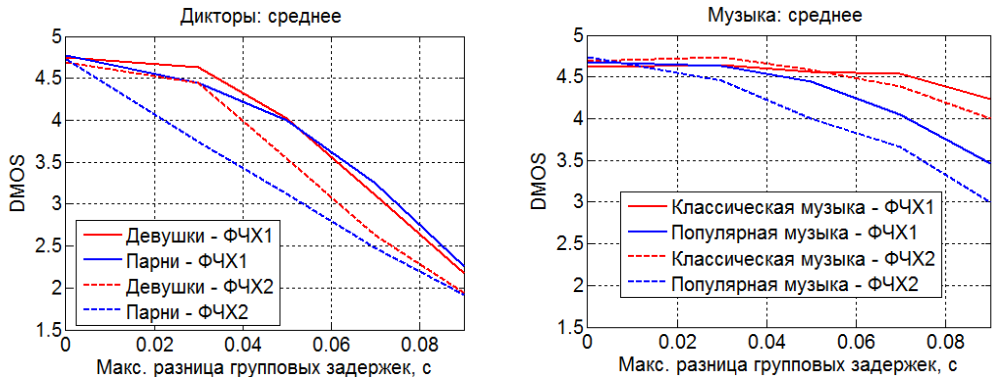
$$BSD = \frac{\sum_{l}^{L}\sum_{k=0}^{\frac{K}{2}-1}\left[B\{X(l,k)\} - B\{Y(l,k)\}\right]^2}{\sum_{l}^{L}\sum_{k=0}^{\frac{K}{2}-1}\left[B\{X(l,k)\}\right]^2}, \quad (3)$$

где $x(l,n)$ и $y(l,n)$ - n -я выборка l-го фрейма входного и выходного сигналов фильтра $x(n)$ и $y(n)$, соответственно; $X(l,k)$ и $Y(l,k)$ – амплитудные спектры l -го фрейма сигналов $x(n)$ и $y(n)$, соответственно; $B\{X(l,k)\}$ и $B\{Y(l,k)\}$ – барк-спектры l-го фрейма сигналов $x(n)$ и $y(n)$, соответственно.

Аналитическое описание показателя PESQ весьма громоздко, его можно найти в [6].

3. Результаты субъективного оценивания качества речевых сигналов

На рис. 3 представлены усредненные, по 32 слушателям, а также по отдельным категориям звуковых файлов, результаты субъективного оценивания качества речи и музыки.



(а) **(б)**

Рис. 3. Графики субъективных оценок качества речи (а) и музыки (б)

Представленные результаты, как видим, хорошо согласуются с предварительными выводами работ [4,5] о том, что фазовые искажения музыкальных сигналов значительно менее заметны на слух, нежели фазовые искажения речевых сигналов. Действительно, при $\Delta\tau_{max} \approx 90$ мс оценки качества речевых сигналов близки к 2 баллам по шкале DMOS (рис. 3,а), тогда как оценки качества музыкальных сигналов остаются сравнительно высокими и заключены в интервале 3-4 балла (рис. 3,б). Более того, подтверждены и уточнены предварительные выводы работ [4,5] о так называемых «пороговых» значениях $\Delta\tau_{max}$ для речи и музыки (50 мс и 70

мс, соответственно), при которых фазовые искажения практически перестают восприниматься слуховой системой человека.

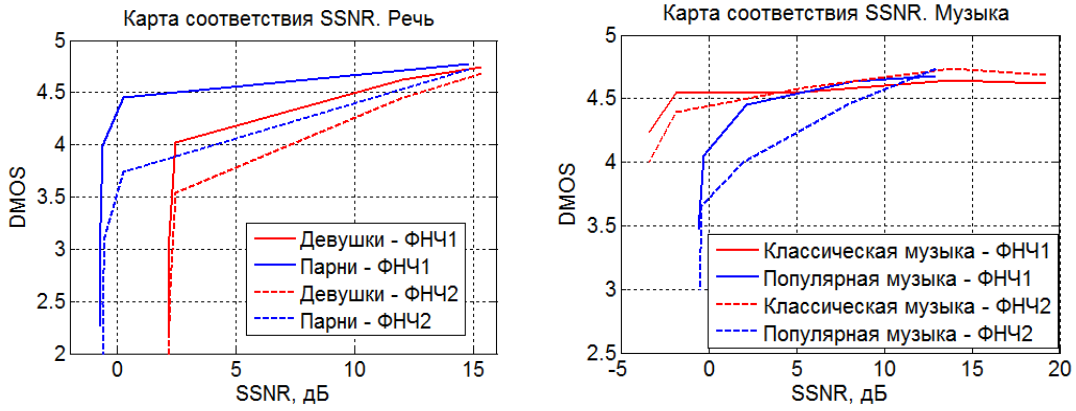
Кроме того, поведение графиков рис. 3,а согласуется с предварительными выводами работы [4] о том, что искажения речевых сигналов для ФЧХ2 заметнее, нежели для для ФЧХ1. Поэтому можно считать подтвержденным приведенное в [4] объяснение, что в ситуации «возрастающее время задержки» при больших $\Delta\tau_{max}$ происходит «превращение» открытых слогов в закрытые. В результате «напевность» речи снижается, вплоть до появления неприятного «дребезга». В ситуации «убывающее время задержки» при больших $\Delta\tau_{max}$ напротив, закрытые слоги «превращаются» в открытые. В результате общее количество открытых слогов растет, и искажения кажутся меньшими за счет увеличения «напевности» речи.

Разумеется, приведенное объяснение не может быть напрямую использовано для аналогичного сопоставления ФЧХ1 и ФЧХ2 при оценивании качества искаженной музыки (рис. 3,б). Тем не менее, некоторая аналогия явно имеет место, поскольку наличие «напевности» у открытых слогов и отсутствие таковой у закрытых слогов является очевидным фактом. Таким образом, приведенные на рис. 3,б результаты свидетельствуют об ошибочности предварительных выводов работы [5] о том, что искажения музыки при ФЧХ1 воспринимаются как более сильные, по сравнению с ФЧХ2. Причиной такой ошибки является, по-видимому, то обстоятельство, что в [5] субъективное оценивание степени и характера искажений музыки осуществлялось единственным экспертом.

4. Сопоставление результатов объективного и субъективного оценивания

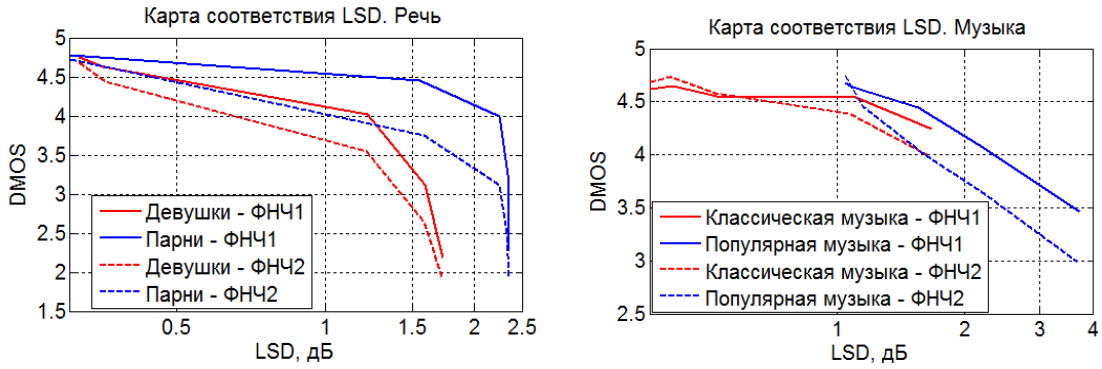
Практическая полезность карт соответствия состоит в возможности пересчета результатов объективного (инструментального) оценивания качества в результаты субъективного оценивания по шкале DMOS.

Карты соответствия объективных и субъективных оценок представлены на рис. 4-7. Как видим, в ряде случаев приведенные зависимости имеют выраженный нелинейный характер, что в определенных ситуациях может значительно затруднять пересчет результатов объективного оценивания в результаты субъективного оценивания по шкале DMOS. Такова, например, ситуация при SSNR ≈ 0 для мужских голосов и при SSNR ≈ 2-2,5 для женских голосов при оценивании качества речи (рис. 4,а). Аналогична ситуация при SSNR ≈ 0 для популярной музыки (рис. 4,б). С классической музыкой дело обстоит несколько лучше, однако объясняется это, главным образом, низкой чувствительностью слуховой системы человека к фазовым искажениям классической музыки. Кроме того, не следует сбрасывать со счетов ограниченность анализировавшегося звукового материала.



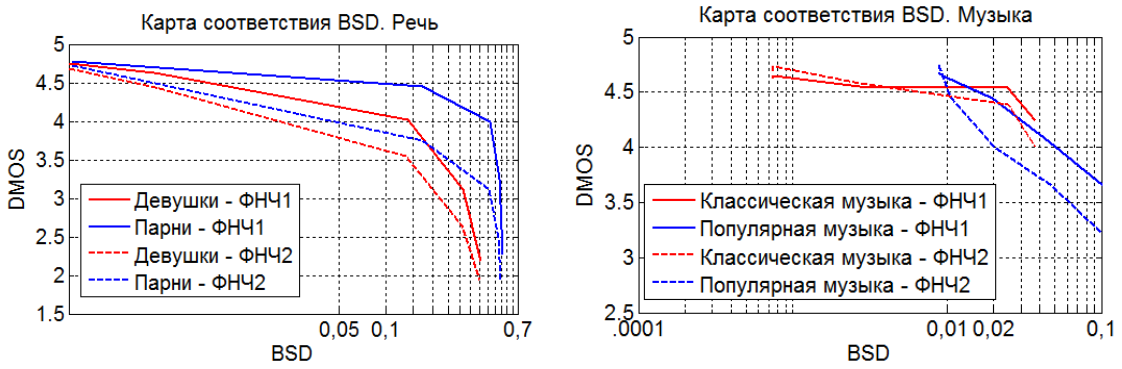
(а) **(б)**

Рис. 4. Графики карт соответствия для меры SSNR: речь (а), музыка (б)



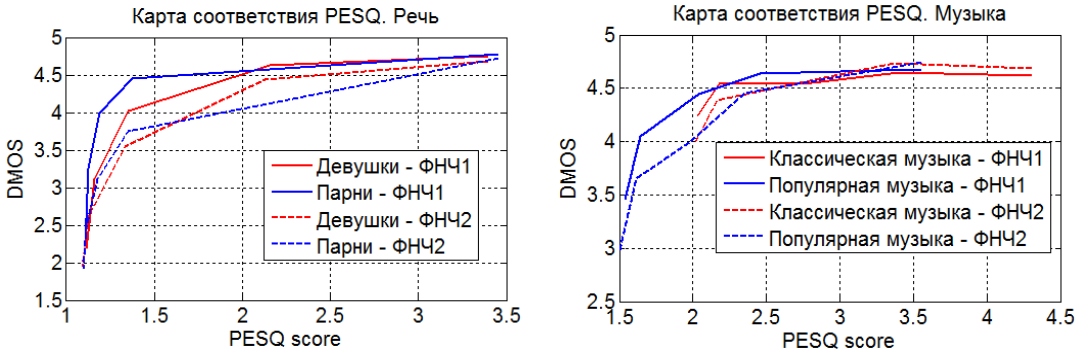
(а) **(б)**

Рис. 5. Графики карт соответствия для меры LSD: речь (а), музыка (б)



(а) **(б)**

Рис. 6. Графики карт соответствия для меры BSD: речь (а), музыка (б)



(а) **(б)**

Рис. 7. Графики карт соответствия для меры PESQ: речь (а), музыка (б)

В силу монотонного характера полученных графиков, отмеченная нелинейность карт соответствия не является препятствием для их практического использования. Хотя, разумеется, точность пересчета объективных оценок в значения шкалы DMOS существенно снижается на участках с повышенной крутизной используемых карт соответствия.

Выводы

Получены усредненные, по 32 слушателям, а также по отдельным категориям звуковых файлов, зависимости результатов субъективного оценивания качества речи и музыки от разницы времен групповой задержки на высоких (8-11 кГц) и низких (90-180 Гц) частотах. При этом уточнены предварительные оценки степени восприятия слуховой системой человека фазовых искажений речевых и музыкальных сигналов, а также подтверждена справедливость предварительного вывода о том, что фазовые искажения музыкальных сигналов значительно менее заметны на слух, нежели фазовые искажения речевых сигналов

Построены карты соответствия результатов объективного и субъективного оценивания качества сигналов. Показано, что в ряде случаев приведенные зависимости имеют выраженный нелинейный характер, что в определенных ситуациях может значительно затруднять пересчет результатов объективного оценивания в результаты субъективного оценивания по шкале DMOS.

Список использованных источников

1. *Дидковский В.С., Дидковская М.В., Продеус А.Н.* Акустическая экспертиза каналов речевой коммуникации. Монография. – К.: Имэкс-ЛТД, 2008. – 420 с.
2. *Martin R., Heute U. and Antweiler C.* (Ed.) Advances in Digital Speech Transmission. – John Wiley & Sons Ltd, England, 2008. – 572 p.
3. *Blauert J.* Group delay distortions in electroacoustical systems / Blauert J. // J. Acoust. Soc. Am., vol.63, No.5, 1978. – P. 1478-1483.
4. *Продеус А.Н., Пилипенко К.П., Калюжный А.Я., Бартенев С.Г.* Оценка влияния нелинейности фазовой частотной характеристики системы на качество речевых сигналов / Электроника и связь, т.20, №2(85), 2015. - С.33-40.
5. *Продеус А.Н., Богданова Н.В.* Оценка влияния нелинейности фазовой частотной характеристики тракта на качество музыкальных сигналов / Electronics and Communications, Vol. 20, No. 4(87), 2015. – P. 29-35.
6. Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.
7. *Cote N.* Integral and diagnostic intrusive prediction of speech. - Springer-Verlag Berlin Heidelberg. – 2011. – 267 p.

Поступила в редакцию 01 июня 2016 г.

УДК 004.934

І.В. Котвицький, **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 і суб'єктивне оцінювання якості мовленнєвих і музичних сигналів, підданих фазовим спотворенням

Недавні попередні дослідження показали, що для слухової системи людини фазові спотворення музичних сигналів є прийнятними, якщо максимальна різниця групових часів затримки тракту в області високих і низьких частот не перевищує 70 мс. Для мовленнєвих сигналів ця величина є меншою й близькою до 50 мс. У даній роботі отримано більш точні суб'єктивні оцінки залежності якості мовлення і музики від різниці часів групової затримки. Крім того, побудовано карти відповідності результатів об'єктивного і суб'єктивного оцінювання якості спотворених сигналів. Показано, що за певних умов такі карти мають виражений нелінійний характер. Бібл. 7, рис. 7, табл. 1.

Ключові слова: *фазові спотворення; суб'єктивна оцінка; якість мовленнєвого сигналу; якість музичного сигналу; показники якості.*

UDC 004.934

I. Kotvytskyi, **A. Prodeus**, Dr.Sc.

National Technical University of Ukraine "Kyiv Polytechnic Institute",

st. Polytechnique, 16, Kiev, 03056, Ukraine.

Objective and subjective evaluation of the quality of speech and music signals subjected to phase distortions

Recent preliminary studies have shown that phase distortion of music signals are acceptable for human auditory system when the maximum difference of group delay time in high and low frequencies is less than 70 ms. This value is less than 50 ms for speech signals. In this paper, we obtain a more accurate subjective speech quality assessment and its dependence on the group delay times difference. In addition, the matching maps for the results of objective and subjective assessment of distorted signals quality were obtained. It has been shown that such maps have a pronounced nonlinear character under certain conditions. Ref. 7, figures 7, table 1.

Keywords: *phase distortion; subjective assessment; speech quality; music quality; quality indicators*.

References

1. *Didovskiy, V., Didovskaia, M. and Prodeus, A.* (2008). Acoustic assessment of speech communication channels. Monograph, K.: Imex-Ltd, P. 420. (Rus)
2. *Martin, R., Heute, U. and Antweiler, C. (Ed.)* (2008). Advances in Digital Speech Transmission, John Wiley & Sons Ltd, England, P. 572.
3. *Blauert, J.* (1978). Group delay distortions in electroacoustical systems. J. Acoust. Soc. Am., vol.63, No.5, Pp. 1478–1483.
4. *Prodeus, A., Pylypenko, K., Kalyuzhny, A. and Bartenev, S.* (2015). Assessment of the impact of system phase response non-linearity on the speech signals quality. Electronics and Communications, Vol.20, No. 2(85), Pp.33–40. (Rus)
5. *Prodeus, A. and Bogdanova, N.* (2015). Evaluation of the effect of phase frequency response non-linearity of the music signals quality. Electronics and Communications, Vol. 20, No. 4(87), P. 29–35. (Rus)
6. Perceptual Evaluation of Speech Quality (PESQ) (2005), ITU-T Recommendations P.862, P.862.1, P. 862.2. Version 2.0.
7. *Cote, N.* (2011). Integral and diagnostic intrusive prediction of speech. Springer-Verlag Berlin Heidelberg, P. 267.