Теория сигналов и систем

УДК 004.934

А.Н. Продеус[1], д.-р.техн.наук, **К.П. Пилипенко**[1], канд.техн.наук,
А.Я. Калюжный[1] д.-р. физ.-мат.наук, **С.Г. Бартенев**[2], канд.мед.наук
[1]Национальный технический университет Украины «Киевский политехнический институт»,
ул. Политехническая 16, 03056, Киев, Украина.
[2] Институт педиатрии, акушерства и геникологии НАМН Украины,
ул. Платона Майбороды 8, г. Киев, 04050, Украина.

Оценка влияния нелинейности фазовой частотной характеристики системы на качество речевых сигналов

Установлено, что для слуховой системы человека приемлемыми являются фазовые искажения речевых сигналов, если максимальная разница групповых времен задержки в области высоких и низких частот не превышает 50 мс – в этом случае интерференция между смежными гласными и согласными звуками на слух практически незаметна. Указаны значения объективных показателей качества речи в виде сегментного отношения сигнал-шум (SSNR), логарифмически-спектральных искажений (LSD), барк-спектральных искажений (BSD) и перцептуальной оценки качества речи (PESQ), соответствующие найденному пороговому значению 50 мс. Библ. 7, рис. 6, табл. 1.

Ключевые слова: *гребенка фильтров; фазовые искажения; качество речевого сигнала; показатели качества.*

Введение

В [5] рассмотрена модель возникновения фазовых искажений сигнала в гребенке цифровых нерекурсивных фильтров с сумматором (рис. 1,а). Данная модель возникновения фазовых искажений сигнала представляет значительный практический интерес, поскольку гребенки фильтров широко используются в системах записи и воспроизведения, кодирования и декодирования акустических сигналов, в линиях связи, в системах коррекции слуха [1]. При значительной нелинейности фазовой частотной характеристики (ФЧХ) такой гребенки искажения выходного сигнала становятся ощутимыми на слух. Введя линии задержки (ЛЗ) в частотные каналы (рис. 1,б), этот недостаток можно устранить. Поскольку использование ЛЗ усложняет систему в целом, возникает естественный вопрос о допустимой степени нелинейности ФЧХ линейной системы при передаче речевых сигналов.

Работ, посвященных исследованию данного вопроса, немного. В [6] отмечено, что удобной мерой нелинейности фазы $\theta(f)$ является отклонение групповой задержки $\tau(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df}$ от константы. Чувствительность слуховой системы человека к неравномерности $\tau(f)$ исследована в [2], где экспериментально показано, что 1-3 мс является пороговым значением, а для тренированного слуха этот порог снижается даже до 400 мкс. Вместе с тем, в [2] указано, что при использовании речевых сигналов фазовые искажения менее заметны. К сожалению, для речевых сигналов пороговые значения неравномерности $\tau(f)$ так и не были определены, хотя для разработчиков электроакустических систем именно эти значения представили бы наибольший интерес.

Другой стороной затронутого вопроса является выбор метода оценивания качества сигнала, а также выбор показателей качества искаженного сигнала. Очевидным и существенным недостатком субъективных методов, использовавшихся в [2], является их высокая ресурсоемкость. Объективные (инструментальные) методы оценивания качества речевого сигнала в значительной степени свободны от указанного недостатка, хотя в литературе отсутствуют четкие рекомендации по выбору таких показателей.

Цель данной работы состоит в восполнении, хотя бы частичном, указанных выше пробелов.

1. Модель возникновения фазовых искажений сигнала

Усложним модель возникновения фазовых искажений сигнала, рассмотренную в [5], увеличив количество октавных фильтров гребенки с пяти до семи и тем самым охватив типичную для речевых сигналов полосу частот 90-11000 Гц. Основные параметры фильтров, состав-

ляющих гребенку и рассчитанных методом Ремеза, приведены в табл. 1, где f_0 - центральная частота; Δf - полоса пропускання; n - порядок фильтра.

Таблица 1. Параметры гребенки октавных фильтров

f_0 , Гц	Δf , Гц	n
125	90	4353
250	180	2903
500	355	2177
1000	710	1320
2000	1400	927
4000	2800	545
8000	5600	437

Амплитудная частотная характеристика (АЧХ) каждого фильтра гребенки близка к прямоугольной, а ФЧХ является линейной в силу симметричности импульсных характеристик (ИХ) каждого из фильтров.

Хотя суммарные АЧХ обеих гребенок одинаковы, ФЧХ гребенки рис. 1,а нелинейная (рис. 1,в), поскольку протяженности парциальных ИХ каждого из полосовых фильтров различны и, как следствие, суммарная ИХ гребенки рис. 1,а несимметрична. Наличие ЛЗ во второй гребенке (рис. 1,б) позволяет добиться симметричности суммарной ИХ и линейности ФЧХ (рис. 1,г). Достигают этого, управляя временами задержки таким образом, чтобы максимумы парциальных ИХ каждого из каналов совпали по времени.

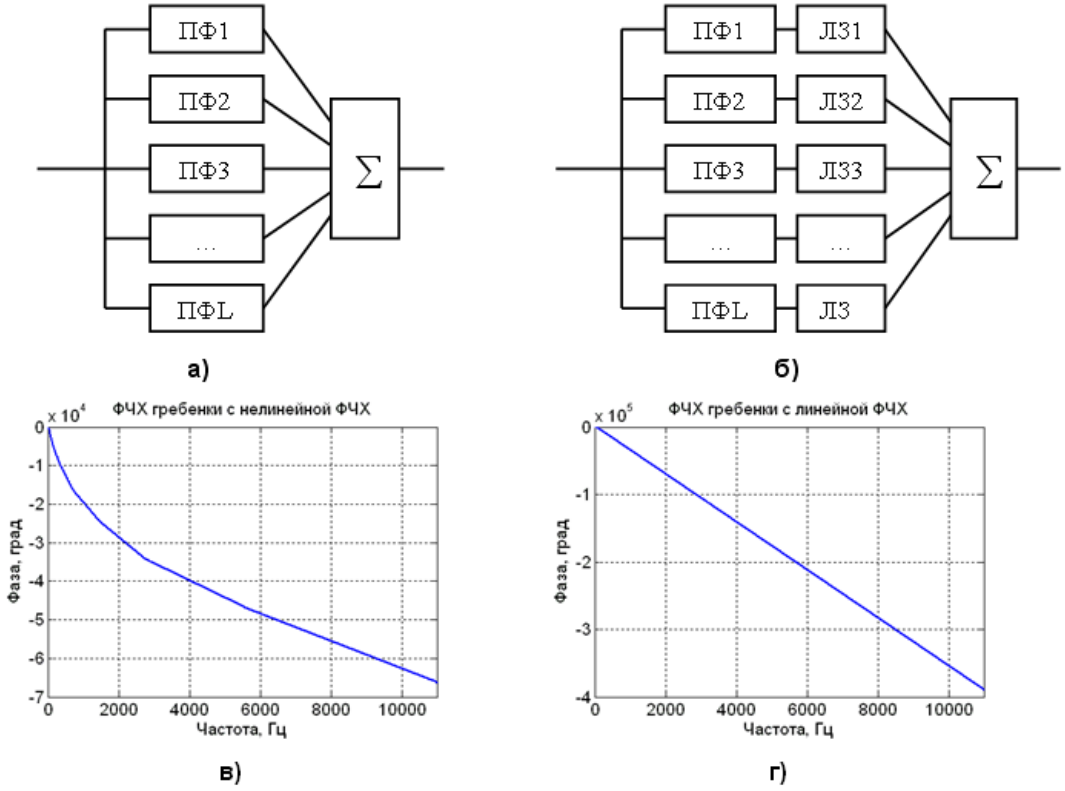
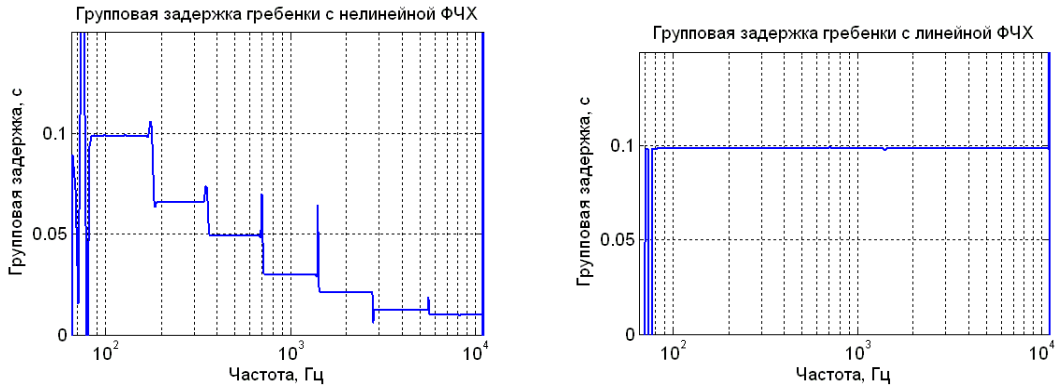


Рис. 1. Изображение гребенок фильтров: без ЛЗ (а) и с ЛЗ (б); ФЧХ гребенок: без ЛЗ (в) и с ЛЗ (г)

Графики группового времени задержки $\tau(f)$, показанные на рис. 2,а, свидетельствуют, что в гребенке с нелинейной ФЧХ (далее - ФЧХ1) различные частотные компоненты сигнала задержаны на время от 10 до 100 мс.

Для гребенки с линейной ФЧХ (далее - ФЧХ2) временная задержка практически постоянна на всех частотах и близка 100 мс (рис. 2,б), поэтому входной и выходной сигналы практически идентичны по форме.



а) **б)**

Рис. 2. Графическое изображение группового времени задержки гребенок фильтров: без ЛЗ (а) и с ЛЗ (б)

Иначе обстоит дело с выходным сигналом гребенки с нелинейной ФЧХ, показанной на рис. 1,в. Максимальное различие времен задержки в данном случае близко $\Delta\tau_{max} \approx 90$ мс, что сопоставимо со среднестатистической протяженностью фонемы (примерно 135 мс). Такая степень нелинейности ФЧХ приводит к искажениям, легко обнаруживаемым на слух и визуально (рис. 3).

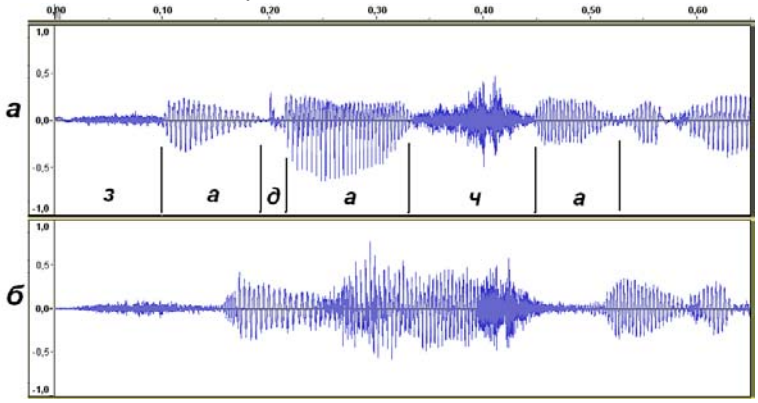
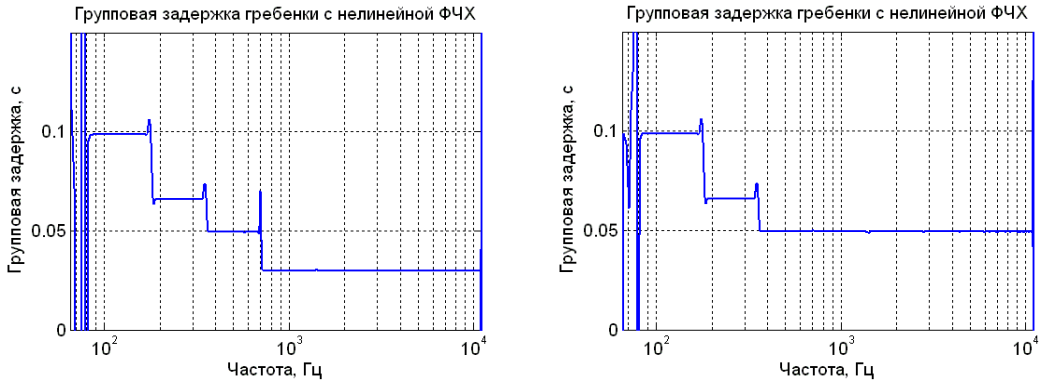
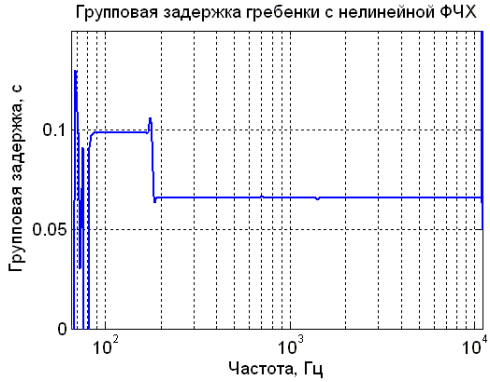


Рис. 3. Графики входного (а) и выходного (б) сигналов для гребенки с ФЧХ1

Наибольший интерес для инженерных приложений представляют «пороговые» значения величины $\Delta\tau_{max}$, при которых искажения речевого сигнала становятся заметными на слух. Кроме того, целесообразно выяснить, каким значениям объективных показателей качества соответствуют выявленные пороговые значения величины $\Delta\tau_{max}$. С этой целью в данной работе, помимо $\Delta\tau_{max} \approx 90$ мс рассмотрены значения $\Delta\tau_{max} \approx 30$, 50 и 70 мс, соответствующие показанным на рис. 4 конфигурациям $\tau(f)$ группового времени задержки.



а) **б)**



в)

Рис. 4. Графики группового времени задержки: $\Delta\tau_{max} \approx 0{,}07$ **с (а),** $\Delta\tau_{max} \approx 0{,}05$ **с (б) и** $\Delta\tau_{max} \approx 0{,}03$ **с (в)**

Такой вид конфигураций $\tau(f)$, где ВЧ компоненты сигнала отстают от НЧ компонентов, в дальнейшем для краткости будем именовать как «убывающее время задержки». Такие же значения $\Delta\tau_{max}$ рассмотрены в данной работе и для ситуации «возрастающее время задержки».

2. Оценивание качества акустического сигнала

Из множества известных на сегодняшний день объективных показателей качества речи рассмотрим четыре: сегментное отношение сигнал-шум (Segmental Signal to Noise Ratio - SSNR), логарифмически-спектральные искажения (Logarithmic Spectral Distortion - LSD), барк-спектральные искажения (Bark Spectral Distortion - BSD) и перцептуальное качество речи (Perceptual Evaluation of Speech Quality - PESQ) [3, 4]. Обосновывая такой выбор, отметим, что первые два показателя – SSNR и LSD – весьма привлекательны в силу простоты вычислений, тогда как другие два показателя – BSD и PESQ – позволяют учесть, с различной степенью точности, особенности слуховой системы человека.

Аналитическое описание упомянутых выше показателей SSNR, LSD и BSD:

$$SSNR = \frac{1}{L}\sum_{l=1}^{L} 10\lg\left[\frac{\sum_{n=Rl}^{Rl+N-1} x^2(l,n)}{\sum_{n=Rl}^{Rl+N-1}[x(l,n)-y(l,n)]^2}\right], \qquad (1)$$

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

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

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

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.

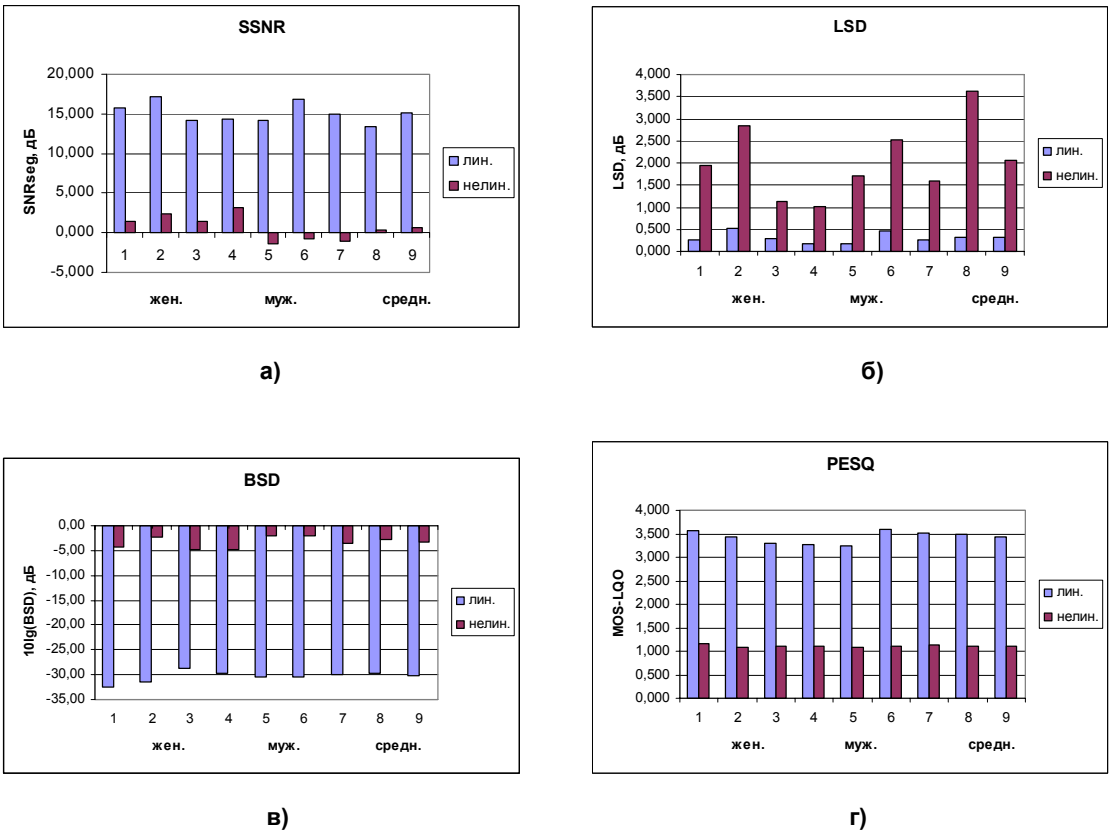
Аналитическое описание алгоритма вычисления показателя PESQ весьма громоздко, его можно найти в [4]. Заметим, что следует учитывать существование двух версий PESQ: ранней (ITU-T Rec. P.862) и поздней (ITU-T Rec. P.862.2). Вторая версия, использованная в данной работе, более совершенна, поскольку позволяет анализировать речевые сигналы в широкой полосе частот (до 7 кГц).

3. Результаты оценивания качества речевых сигналов

При экспериментальном оценивании, как субъективном, так и объективном, зависимости качества сигнала от степени нелинейности ФЧХ, использованы фрагменты, протяженностью 1 минута каждый, речевых сигналов для 4-х дикторов-женщин и 4-х дикторов-мужчин, читающих русский текст по юридической темати-

ке. Запись сигналов произведена на кафедре акустики НТУУ «КПИ», в заглушенном помещении с временем реверберации 0,15 с, с частотой дискретизации 22050 Гц и битовой глубиной 16 бит.

Оценки объективных показателей для речевых сигналов, пропущенных через первую ($\Delta\tau_{max} \approx 90$ мс) и вторую ($\Delta\tau_{max} \approx 0$ мс) гребенки, приведены на рис. 5, где первые четыре пары столбцов соответствуют дикторам-женщинам, вторые четыре пары – дикторам-мужчинам, последняя пара столбцов представляет средние по всем дикторам результаты. Как видим, все рассмотренные показатели качества адекватно отреагировали на нелинейность ФЧХ гребенки фильтров, засвидетельствовав существенное ухудшение качества речевого сигнала при $\Delta\tau_{max} \approx 90$ мс. При этом результаты для «убывающего» и «возрастающего» времени задержки практически совпадают.



а) **б)**

в) **г)**

Рис. 5. Значения объективных показателей

Варьирование значениями $\Delta\tau_{max}$ позволило субъективно оценить пороговое значение чувствительности слуховой системы человека к фазовым искажениям речевого сигнала. Анализ ситуации «убывающее время задержки» показал, что на слух первые признаки искажений в виде небольшой «сиплости» звучания речи появляются при $\Delta\tau_{max} \approx 50$ мс. Дальнейшее увеличение $\Delta\tau_{max}$ до 70-90 мс приводит к эффекту «сиплый хорус», при котором речевой сигнал воспринимается как одновременное чтение текста несколькими дикторами со слегка осипшими голосами.

Зависимости усредненных (по дикторам) оценок объективных показателей качества от $\Delta\tau_{max}$ представлены на рис. 6, где видно, что при $\Delta\tau_{max} \approx 0{,}03$ с гребенка с нелинейной ФЧХ в наименьшей степени проигрывает гребенке с линейной ФЧХ. С переходом от значения $\Delta\tau_{max} \approx 0{,}03$ с к субъективно определенному «пороговому» значению $\Delta\tau_{max} \approx 0{,}05$ с, скорость ухудшения качества гребенки с нелинейной ФЧХ максимальна, а затем, с ростом $\Delta\tau_{max}$, эта скорость уменьшается.

Анализ ситуации «возрастающее время задержки», при которой ВЧ компоненты сигнала задерживаются сильнее НЧ компонентов, показал, что пороговое значение максимального различия времен задержки осталось прежним, т.е. близким 50 мс. Однако увеличение $\Delta\tau_{max}$ до 70-90 мс приводит к заметному изменению характера речи: теперь она носит не сиплый, а «дребезжащий» характер. Субъективно это воспринимается как более сильные, по сравне-

нию с сиплостью, искажения речи. Заметим, что среди объективных показателей лишь оценки PESQ согласовались с таким субъективным впечатлением, оценки остальных показателей дали противоположный результат. Заметим, впрочем, что значения всех оценок объективных показателей отличались лишь на 1-2% для ситуаций «убывающее» и «возрастающее» время задержки.

Как следует из рис. 6, пороговому значению $\Delta\tau_{max} \approx 50$ мс соответствуют следующие пороговые значения объективных показателей: 1 дБ для SSNR; 1,8 дБ для LSD; 0,3 для BSD и 1,4 MOS для PESQ.

Объяснить различие субъективных ощущений, обусловленное переходом от ситуации «убывающее время задержки» к ситуации «возрастающее время задержки» можно следующим образом. Ситуация «убывающее время задержки» негативным образом сказывается на звучании закрытых слогов вида «ас», «ак» и т.п., поскольку в этом случае происходит интерференция существенно сдвинувшегося по оси времени низкочастотного гласного звука с мало сдвинувшимся по времени согласным звуком, в спектре которого важную роль играют высокочастотные компоненты. В ситуации «возрастающее время задержки» наибольшим разрушениям подвергаются открытые слоги вида «са», «ка» и т.п. В этом случае существенно сдвинувшийся по оси времени согласный звук интерферирует со слабо сдвинувшимся гласным звуком.

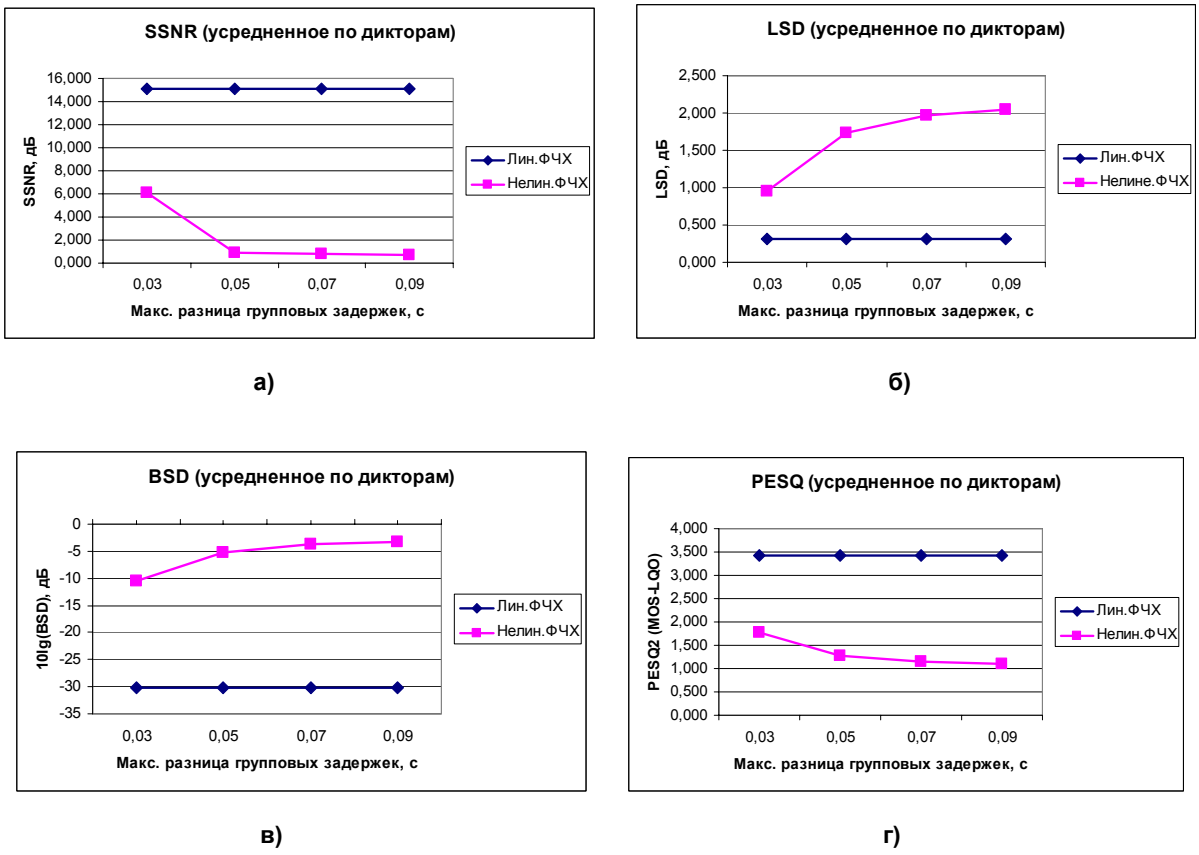


Рис. 6. Графики зависимости объективных показателей от $\Delta\tau_{max}$

Открытые слоги, сообщающие речи «напевность», в русской речи встречаются примерно в 1,5 раза чаще закрытых слогов [7]. В ситуации «возрастающее время задержки» при больших $\Delta\tau_{max}$ происходит «превращение» открытых слогов в закрытые, «напевность» речи снижается, вплоть до появления «дребезга».

В ситуации «убывающее время задержки» при больших $\Delta\tau_{max}$ происходит иное явление: закрытые слоги «превращаются» в открытые, общее количество открытых слогов растет, искажения кажутся меньшими.

Выводы

Экспериментальные исследования показали, что для слуховой системы человека приемлемыми являются фазовые искажения речевых сигналов, если максимальная разница группо-

вых времен задержки в области высоких и низких частот не превышает 50 мс. Объяснить это можно тем, что в этом случае интерференция между смежными гласными и согласными звуками на слух малозаметна.

Найдены пороговые значения объективных показателей, соответствующие пороговому значению $\Delta\tau_{max} \approx 50$ мс: это 1 дБ для SSNR, 1,8 дБ для LSD, 0,3 для BSD и 1,4 MOS для PESQ.

Полученные результаты будут полезными для разработчиков электроакустических систем, поскольку позволяют обосновать требования к фазовым характеристикам аппаратного и программного обеспечения.

Список использованных источников

1. Advances in Digital Speech Transmission / Edited by Martin R., Heute U. and Antweiler C. - John Wiley & Sons Ltd, England, 2008. – 572 p.
2. *Blauert J.* Group delay distortions in electroacoustical systems / Blauert J. // J. Acoust. Soc. Am. – Vol.63, No.5. – 1978. – P. 1478-1483.
3. *Habets E.A.P.* Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement. – PhD dissertation, Eindhoven, 2007. – 257 p.
4. Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.
5. *Дидковский В.С., Дидковская М.В., Продеус А.Н.* Акустическая экспертиза каналов речевой коммуникации. Монография. – К.: Имэкс-ЛТД, 2008. – 420 с.
6. *Оппенгейм А., Шафер Р.* Цифровая обработка сигналов. – М.: Техносфера, 2006. – 858 с.
7. *Смирнова Н. С., Чистиков П. Г.* Программа анализа фонетических статистик в текстах на русском языке и ее использование для решения прикладных задач в области речевых технологий // Матер. XXVII Междунар. конф. «Диалог». – М., 2011. – С. 632-644.

Поступила в редакцию 11 ноября 2014 г.

УДК 004.934

А.М. Продеус[1], д.-р. техн. наук, **К.П. Пилипенко**[1], канд. техн. наук,
О.Я. Калюжний[1] д.-р. фіз.-мат. наук, **С.Г. Бартєнєв**[2], канд. мед. наук
[1]Національний технічний університет України «Київський політехнічний інститут»,
вул. Політехнічна, 16, корпус 12, м. Київ, 03056, Україна.
[2]Інститут педіатрії, акушерства та гінекології НАМН України,
вул. Платона Майбороди 8, м. Київ, 04050, Україна.

Оцінка впливу нелінійності фазової частотної характеристики системи на якість мовленнєвих сигналів

Показано, що для слухової системи людини прийнятними є фазові спотворення мовленнєвих сигналів, якщо максимальна різниця групових часів затримки в області високих і низьких частот не перевищує 50 мс - при такій різниці групових часів затримки інтерференція між суміжними голосними й приголосними звуками є практично непомітною на слух. Вказано значення об'єктивних показників якості мовлення у вигляді сегментного відношення сигнал-шум (SSNR), логарифмічно-спектральних спотворень (LSD), барк-спектральних спотворень (BSD) і перцептуальної оцінки якості мовлення (PESQ), що відповідають знайденому граничному значенню 50 мс. Бібл. 7, рис. 6, табл. 1.

Ключові слова: *гребінка фільтрів; фазові спотворення; якість мовленнєвого сигналу; показники якості.*

UDC 004.934

A.M. Prodeus[1], Dr.Sc., **K.P. Pylypenko**[1], Ph.D., **A.Ya. Kalyuzhny**[1], Dr.Sc., **S.G. Bartenev**[2], Ph.D.
[1]National Technical University of Ukraine "Kyiv Polytechnic Institute",
off. 233, Politekhnichna Str., 16, Kyiv, 03056, Ukraine.
[2] Institute of Pediatrics, Obstetrics and Gynecology of NAMS of Ukraine,
str. Platona Mayborody 8, Kiev, 04050, Ukraine.

Assessment of the impact of system phase response non-linearity on the speech signals quality

It is shown that phase distortion of speech signals are acceptable for human auditory system when the maximum difference of group delay times in the high and low frequencies is below 50 ms - the interference between adjacent vowels and consonants is not perceived with such a difference of group delay. There were founded values of objective measures of speech quality in the form of a segmental signal-to-noise ratio (SSNR), the log-spectral distortion (LSD), bark spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ), according to the detected threshold value of 50 ms. Bibl. 7, Fig. 6, Tab. 1.

Keywords: *filter bank; phase distortion; speech quality; quality indicators.*

References

1. Edited by Martin R., Heute U. and Antweiler C. (2008), Advances in Digital Speech Transmission. John Wiley & Sons Ltd, England, P. 572.
2. Blauert J. (1978), Group delay distortions in electroacoustical systems. J. Acoust. Soc. Am. Vol.63, No.5. Pp. 1478-1483.
3. Habets E.A.P. (2007), Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement. PhD dissertation, Eindhoven, P. 257.
4. Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0. October 2005.
5. Didovskiy V.S., Didovskaia M.V., Prodeus A.N. (2008), "Acoustic assessment of speech communication channels. Monograph," K.: Imex-Ltd, P. 420. (Rus)
6. Oppenheim A., Schafer R. (2006), "Digital signal processing," M.: Techospera, P. 858. (Rus)
7. Smirnova N.S., Chistikov P.G. (2011), "Phonetic analysis program in statistics in Russian texts and its use for applications in the field of speech technology," Proc. XXVII Intern. Conf. «Dialog», M., Pp. 632-644 (Rus)