

АНАЛИЗ ПАРАМЕТРОВ МОДУЛЯЦИИ ИНФОРМАЦИОННЫХ СВЕРХНИЗКОЧАСТОТНЫХ СОСТАВЛЯЮЩИХ (КИНЕМ) РЕЧЕВОГО СИГНАЛА

Проведено аналіз інформаційних параметрів модуляції зверхнизькочастотних складових (кінем) мовного сигналу для шепоту. Експериментально доведено, що в рівноартикуляційних смугах найбільший індекс куттової модуляції, який обумовлює стохастичність амплітудного спектру потужності фонем, є суттєво більшим за одиницю.

Проведен анализ информационных параметров модуляции сверхнизкочастотных составляющих (кинем) речевого сигнала для шепотной речи. Экспериментально доказано, что в равноартикуляционных полосах максимальный индекс угловой модуляции, который определяет стохастичность амплитудного спектра мощности фонем, существенно больше единицы.

The analysis of informational parameters of modulation of the super-low-frequency speech signal components (kinemas) for whisper speech is realized. It is proved experimentally that in the equiarticulation bands the maximum index of angle modulation that determines the stochasticity of phoneme power amplitude spectrum is much more than one.

Введение. Постановка задачи

В соответствии с законом Украины [1], целью информационной защиты является предотвращение утечки, хищения, утраты, искажения и подделки (имитации) информации. С точки зрения информационной разведзащищенности необходимо акцентировать внимание на информационной составляющей речевого сигнала (РС), которая должна быть сокрыта сигналом маскирования (СМ) в точке несанкционированного доступа технической разведки противника (ТРП).

В соответствии с существующими теориями речеобразования и слуха РС рассматривается и анализируется как информационно связанная совокупность дифференциальных составляющих речи – фонем [2]. Однако данный подход противоречит физиологическому процессу речевого обмена – передаче и приему сведений об образах окружающей действительности, которые передаются, анализируются и идентифицируются словами. Для анализа данного противоречия, которое можно характеризовать как проблему теорий речеобразования и слуха, необходимо провести спектрально-модуляционные исследования информационных составляющих слов – кинем, которые позволят идентифицировать информационные составляющие РС на интервале длительности слова.

Задачей статьи является анализ параметров индекса угловой модуляции и спектральной

плотности мощности информационных сверхнизкочастотных составляющих РС, которые обусловлены движениями активных физиологических артикуляционных органов речеобразующей системы.

Основная часть

Источником акустического РС является артикуляционный аппарат диктора. Он состоит из следующих физиологических органов: бронхов, легких, диафрагмы, трахеи, голосовых связок, гортани, глотки, небной занавески, языка, ротовой и носовой полостей. В процессе синтеза РС воздух, под действием диафрагмы, из легких проходит через трахею, голосовые связки, гортань и затем разветвляется на два потока (рис. 1). Один поток поступает в носоглотку, взаимодействует с носовой полостью и выходит через носовые отверстия. Другой поток, пройдя через ротовую полость, выходит через ротовое отверстие. Воздух, сжатый в легких мышцами грудной клетки, поступает в трахею, перекрытую голосовыми связками, которые представляют собой эластичные мышцы. Под действием потока воздуха и управлением параметра эластичности связки меняют свою форму и площадь отверстия голосовой щели. При синтезе вокализованных звуков голосовые связки вибрируют, создавая, сигнал основного тона (ОТ).

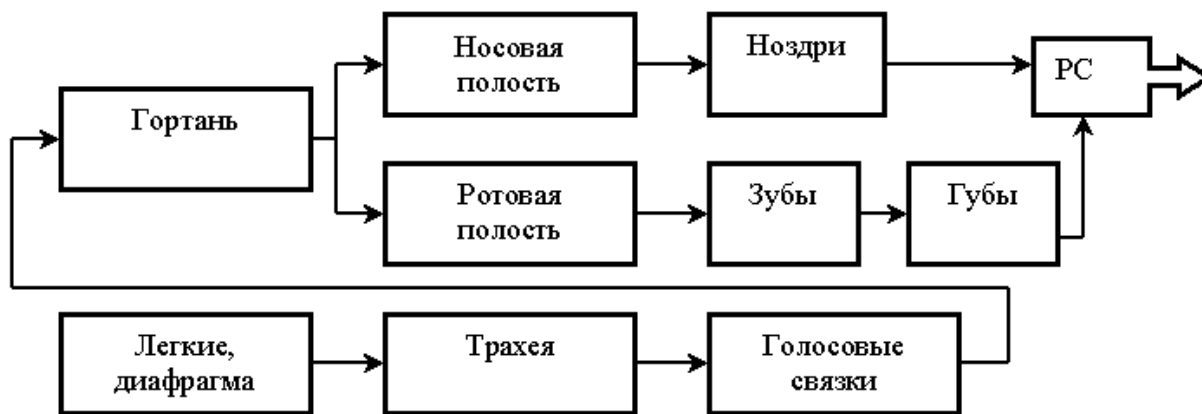


Рис. 1. Функциональная модель артикуляционного аппарата

Невокализованные звуки синтезируются при открытых и неподвижных голосовых связках, под действием физиологического турбулентного шума $S^n(t)$, возникающего в результате прохождения воздушного потока через щели, образуемые артикуляционными физиологическими органами. Сигнал ОТ и турбулентный шум являются несущими сигналами для информационных составляющих РС в воздушной среде. Изменения конфигурации голосового тракта и колебания голосовых связок взаимосвязаны так, что вся артикуляционная система функционирует как единый сложный объект. Одна группа органов – зубы, твердое небо, задняя стенка глотки и носовой полости – участвует в артикуляции пассивно, так как остается неподвижной. Ее можно определить как идентификационную группу артикуляционных органов диктора, в части анализа информации, содержащейся в РС. Другая группа артикуляционных органов – легкие, диафрагма, голосовые связки, нижняя челюсть, губы, язык, мягкое небо, нёбная занавеска, – является активно-информационной, так как в процессе артикуляции осуществляет упорядоченные центральной нервной системой информативные движения. Эти движения физиологических артикуляционных органов русский основатель Казанской лингвистической школы И. А. Бодуэн де Куртенэ объединил [3] общим термином «кинема».

Информационный сигнал кинем $Ki(t)$ является свехнизкочастотным ($f_k < 300\text{Hz}$) и не может быть передан в атмосфере на жизненно необходимые расстояния (за исключением метода сурдопередачи), поэтому в процессе физиологического процесса речеобразования он модулирует несущие сигналы, образованные гармоническими составляющими сигнала ОТ

(процесс вокализации) и физиологического шума, обеспечивая необходимую акустическую мощность РС. К настоящему времени свойства и параметры вокализованного РС рассмотрены и исследованы достаточно подробно [4]. Акцентирование процесса вокализации обосновало вывод о доминирующем информационном значении амплитудной модуляции несущих спектральных составляющих сигнала ОТ, и, соответственно, информационном значении амплитудного спектра плотности мощности РС. Однако форма и параметры сигнала ОТ существенно переменны, потому амплитудный спектр мощности для одних и тех же фонем (слов) отличаются для одного и того же аудитора. Данный факт не позволяет провести модуляционный анализ информационных параметров кинем для вокализованной речи. С целью информационного анализа модуляционных параметров кинем, на данном этапе анализа считаем целесообразным предварительно провести исследования слов, артикулированных «шепотной речью», т.е. РС, в котором отсутствует сигнал ОТ, однако полностью сохраняется информационная составляющая.

Принимая во внимание факт объединения частотных составляющих речи в частотные (равноартикуляционные) группы [5], речевой сигнал $Si(t)$ можно представить как сумму несущих сигналов частотных групп $S^n_{Fi}(t)$, которые модулированы $M_{A,\phi}[S^n_{Fi}(t), Ki_{Fi}(t)]$ информационными составляющими кинем соответствующих частотных групп $Ki_{Fi}(t)$:

$$Si(t) = \sum_{i=1}^k M_{A,\phi}[S^n_{Fi}(t), Ki_{Fi}(t)], \quad (1)$$

где k – количество анализируемых частотных групп РС.

Основываясь на гипотезе фазовой чувствительности рецептора слуха [6], представим сигнал кинем $Ki(t)$ как сумму m гармонических сигналов

$$Ki(t) = \sum_{i=1}^m A_i(t) \cos[\omega k_i t + \beta_i(t) \Psi_i(t)],$$

где выражение под знаком косинуса есть полная фаза, которая является информационным сигналом, т.к. отображает информацию о текущем изменении фазы сигнала кинем. Несущий сигнал физиологического шума $S^n(t)$ классически [4] представляет собой случайный процесс с нормальным законом распределения плотности вероятности амплитуд. Таким образом, сигналы (1) слов шепотной речи можно представить следующим выражением:

$$Si(\omega, t) = S^n(\omega^n, t) \times \sum_{i=1}^m A_i(t) \cos[\omega k_i t + \beta_i(t) \Psi_i(t)], \quad (2)$$

где $\Psi_i(t)$, ωk_i – фазовый угол и угловая частота i -ой кинемы, $\beta_i(t)$ – соответствующий индекс угловой модуляции.

Данный сигнал можно анализировать как широкополосный случайный процесс в полосе несущих тональных частот: ω^l, ω^h – нижняя и верхняя частоты анализируемого РС.

Детерминированной характеристикой данного процесса (2) является средняя мощность $N(\omega, t)$, которая определяется [7] как преобразование Фурье от его автокорреляционной функции $R_{Si}(\tau)$:

$$N(\omega^n, t) = \frac{1}{\pi} \int_0^{Tc} R_{Si}(\tau) \exp(-j\omega^n \tau) d\tau, \quad (\omega^n \in [\omega^l, \omega^h]; \tau \in [0, Tc]), \quad (3)$$

где Tc – анализируемый временной интервал длительности слова.

Однако преобразование Фурье имеет локализацию в частотной области, т.к. его ядро $\exp(-j\omega^n \tau)$ не локализовано во времени τ . В связи с тем, что процесс артикуляции кинем иррегулирован во времени, локализуем преобразование Фурье на временных промежутках конечной длины, которая определяется нижней частотой кинем ωk^l , вводя в процесс преобразования Гэбора [8] окно Гаусса $W_G(\omega^n, T_G)$

с временным интервалом $T_G = \frac{2\pi}{\omega k^l}$, подставляя в (3), получим:

$$N(\omega^n, t) = \frac{1}{\pi} \int_0^{Tc} W_G(\omega^n, T_G) R_{Si}(\tau) \exp(-j\omega^n \tau) d\tau, \quad (\omega^n \in [\omega^l, \omega^h]; \tau \in [0, Tc]) . \quad (4)$$

Экспериментальные исследования проводились с реальными РС, артикулированными из стандартных словесных таблиц профессиональным диктором и сигналами физиологического шума. В качестве демонстрационного компонента РС для данной статьи взято слово «лошадка», в связи с тем, что оно содержит гласные, согласные и взрывные фонемы. Длительность слова 884 мс (8123 отсчетов при частоте дискретизации 44100 Гц).

В первую очередь, с целью корректности применения метода [7], проводился анализ параметров нормальности физиологического шума $S^n(t)$. Результаты анализа (на интервалах времени $T_a = 10$ ms, $\omega^l = 100$ Hz) распределения амплитуд, спектра и автокорреляционной функции позволяют сделать вывод о нормальном законе распределения плотности вероятности амплитуд, и, таким образом о корректности применения метода идентификации огибающей и фазы случайного сигнала.

Анализ спектральной плотности мощности $N(\omega^n, t)$ по выражению (4) проводился в подтональном диапазоне частот $\omega^n < 300$ Hz с временным интервалом окна анализа $T_G = 500$ ms, $\omega k^l = 2$ Hz, которое перемещалось по данным файла анализируемого слова дискретно с шагом $T_d = 11$ ms. Результаты анализа приведены на рис. 2.

На графике зависимости спектральной плотности мощности от частоты и времени видны максимумы в районе $f_k = (12; 25; 125)$ Hz, которые кратны частотам девиации формант [10], что подтверждает корректность полученных результатов. Спектральная составляющая с частотой сети $f = (50)$ Hz является помехой, в связи с тем, что при записи слов применялся аналоговый микрофон. В процессе эксперимента был проведен анализ 200 слов (4 таблицы по 50 слов). Проведя анализ полученных результатов, можно сделать вывод о реальности информационной угловой модуляции кинемами несущих частот физиологического шума.

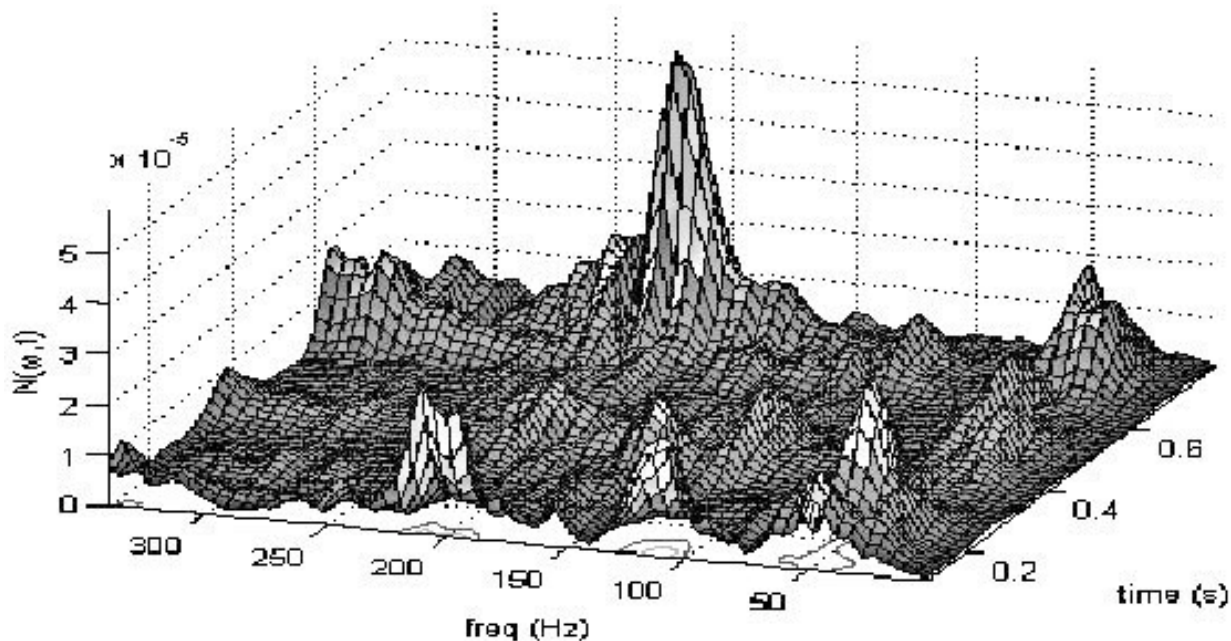


Рис. 2. Результаты анализа временной и частотной зависимости спектральной плотности мощности $N(\omega^n, t)$ слова «лошадка».

Для проведения анализа индекса угловой модуляции кинемами несущих сигналов в частотных (равноартикуляционных) группах [5] сделаем допущение, что на интервале времени анализа T_a огибающая спектра мощности сигнала кинемы, которая является параметром адаптации коэффициента передачи среднего уха, постоянна и равна единице. В этом случае информационная составляющая РС, анализируемая слуховым рецептором, будет определяться девиацией параметра индекса угловой модуляции $\beta(t)$. Демодуляция РС осуществлялась традиционным методом [8], который предусматривает промежуточное вычисление аналитического сигнала при помощи преобразования Гильберта и нахождение аргумента $\Psi_i(\omega_i^n, t)$ исследуемого сигнала.

$$\Psi_i(\omega_i^n, t) = \arg \{ Hil \{ S_i^n(\omega_i^n, t) \times \cos[\omega_k t + \beta_i(t) \Psi_i(t)] \exp(-j\omega_i^n t) \} \}, \quad (5)$$

где $Hil\{*\}$ – вычисление преобразования Гильберта, $\arg\{*\}$ – вычисление аргумента функции.

В связи с тем, что при модуляции кинемами $Ki(t)$ нормального физиологического шума $S^n(t)$ несущая частота ω_i^n информационно флуктуирует возле своего среднего значения, анализ ее девиации осуществлялся оригинальным методом, предусматривающим итерационное вычисление несущих частот кинем (5) по критериям стационарности исследуемого демо-

дулированного аналитического сигнала (постоянства математического ожидания $M[*]$ и дисперсии $D[*]$ с задаваемым диапазоном точности $\pm x$):

$$\left\{ \begin{aligned} M_m[S_{i_{Fi}}(\omega_i, t)] &= M_d[S_{i_{Fi}}(\omega_i, t)] \pm x_{M,d} \\ D_m[S_{i_{Fi}}(\omega_i, t)] &= D_d[S_{i_{Fi}}(\omega_i, t)] \pm x_{D,d} \end{aligned} \right\},$$

где подстрочные прописные индексы m, d относятся к соответствующим процессам модуляции и демодуляции, x_d – параметр допустимого отклонения при демодуляции определялся как $x_{M,d} = 0,05M_m$.

Индекс угловой модуляции в выражении (5) определялся для аналитического сигнала как отношение максимальной девиации аргумента к максимальной девиации несущего сигнала

$$\beta(t) = \frac{\Psi_i^{\max}(\omega_i^n, t) - \Psi_i^{\min}(\omega_i^n, t)}{\pi},$$

где $\Psi_i^{\max}(\omega_i^n, t), \Psi_i^{\min}(\omega_i^n, t)$ – максимальное и минимальное значение аргумента выражения (5) для i -ой частотной полосы. Адекватность метода подтверждена инверсным восстановлением анализируемого слова.

Данные слов анализировались на усредненных временных интервалах фонем (bloc) $T_a = 94 \text{ ms}$ в 19 равноартикуляционных частотных полосах [5] (band). Результаты анализа параметра индекса угловой модуляции $\beta(t)$ слова «лошадка» приведены на рис. 3.

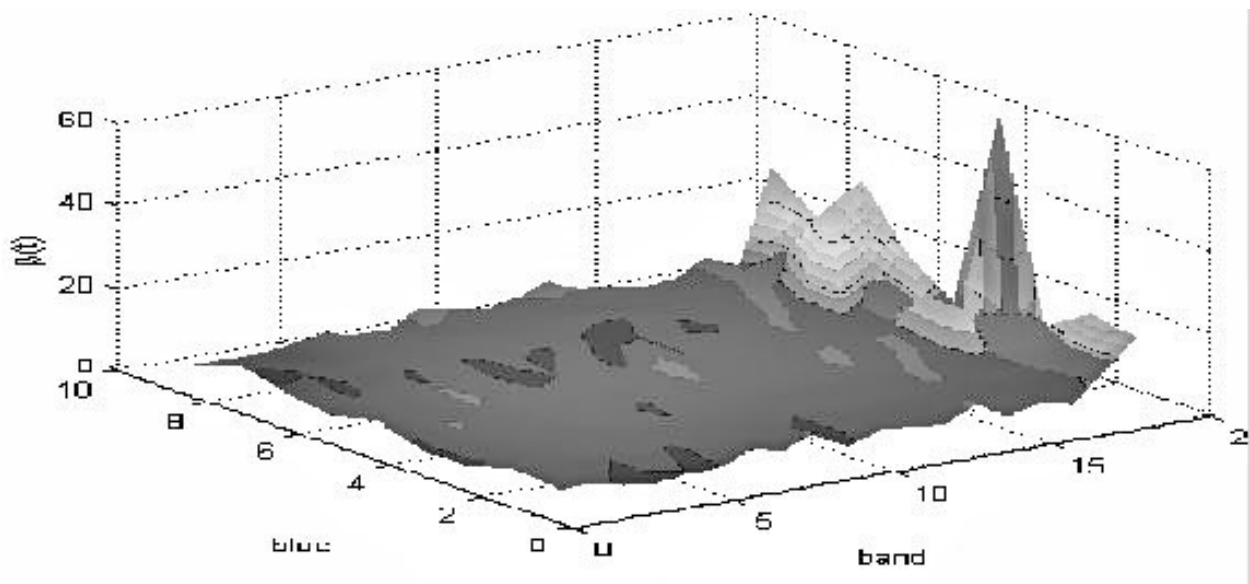


Рис. 3. Результаты анализа зависимости информационного параметра индекса угловой модуляции $\beta(t)$ от частоты (band) и времени (bloc) слова «лошадка», артикулированного шепотной речью.

Максимальные значения параметра $\beta(t)$ наблюдаются в полосе частот первых трех формант, а также высокочастотной области спектра, что хорошо согласуется с результатами ранее проведенных исследований [2, 4].

Для сравнения максимального индекса модуляции были проведены испытания со

словами, которые артикулированы вокализованной речью, в частности на рис. 4 представлена зависимость параметра индекса угловой модуляции $\beta(t)$ от частоты (band) и времени (bloc) слова «лошадка» (длительность 1017 мс), артикулированного вокализованной речью.

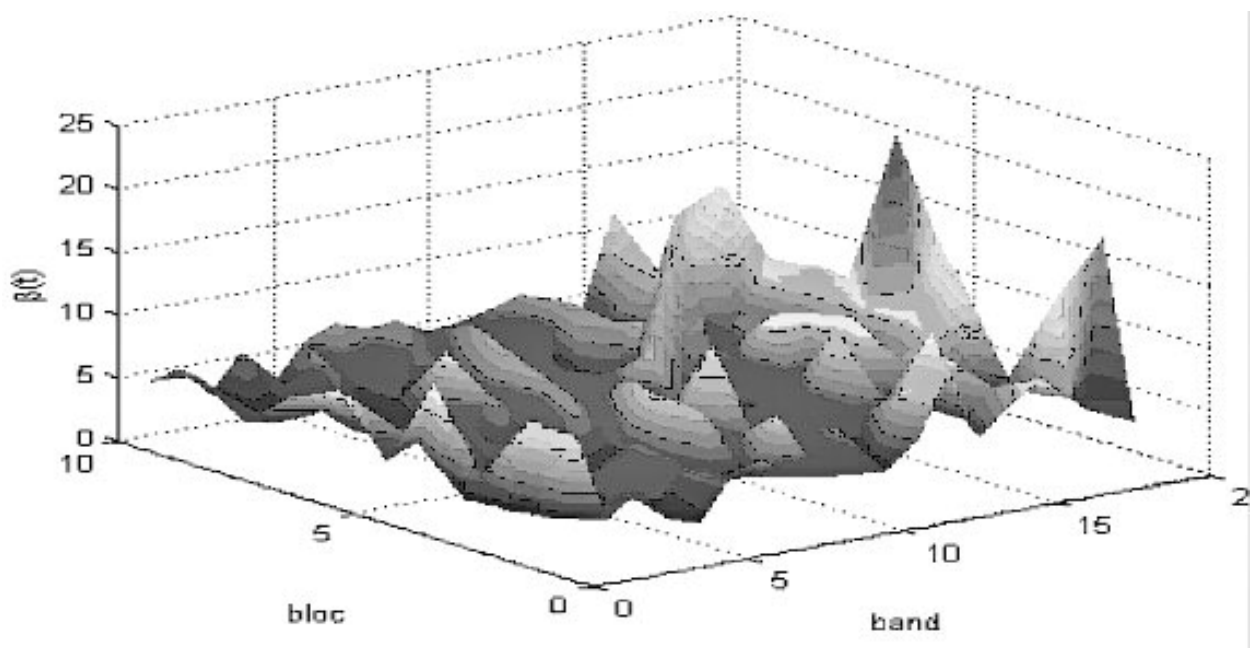


Рис. 4. Результаты анализа зависимости информационного параметра индекса угловой модуляции $\beta(t)$ от частоты (band) и времени (bloc) слова «лошадка», артикулированного вокализованной речью.

Сравнивая рис. 3 и рис. 4, видно, что значение параметра индекса угловой модуляции $\beta(t)$ претерпело значительные изменения: появились максимумы на частотах первых трех фор-

мант, а также значительно уменьшилась амплитуда максимума на частотах пятой форманты. Исходя из анализа индекса модуляции, можно сделать вывод, что спектр сигнала одного и то-

го же слова также значительно изменился. Для выяснения причины столь значительной вариативности спектра проведем анализ ширины спектра в равноартикуляционных полосах. Известно [9], что спектр $S_{\varphi}(\omega, t)$ сигнала с угловой модуляцией определяется как:

$$S_{\varphi}(\omega, t) = A(t) \sum_{i=-\infty}^{i=\infty} J_i[\beta(t)] \cos(\omega_i^n + i\omega k_i)t, \quad (6)$$

где $J_i[\beta(t)]$ – функция Бесселя 1-го рода порядка i от аргумента $\beta(t)$. Как видно из анализа выражения (6), спектр РС в каждой равноартикуляционной полосе чувствителен к параметру $\beta(t)$ и содержит бесконечное количество составляющих с частотами $\omega_i^n + i\omega k_i$; $k = 0, \pm 1, \pm 2, \dots$. Амплитуда i -й составляющей равна $A(t)J_i[\beta(t)]$, т.е. пропорциональна функции Бесселя i -ого порядка от аргумента $\beta(t)$. При индексах модуляции, больших единицы [10], ширина спектра $S_{\varphi}(\omega, t)$ близка к удвоенной девиации частоты кинем $2\omega k$, на границах равноартикуляционных полос происходит суперпозиция спектров сигналов, что и определяет значительную вариативность и неравномерность [11] спектра фонем РС.

Выводы

В процессе анализа исследовались частоты кинем $Ki_F(t)$, частоты девиации несущих частот частотных групп на интервале времени длительности слова, а также максимальные индексы угловой модуляции несущих частот в полосе частотной группы. Исследования проводились со словами, артикулированными как шепотной, так и вокализованной речью.

В результате исследований установлено:

1) Частоты кинем $\omega k_i(t)$ лежат в диапазоне (2...250) Гц, с максимумом (24...30, 60...100) Гц, что хорошо согласуется с нечетными гармониками частоты девиации первых трех формант.

2) Для шепотной речи индекс угловой модуляции $\beta(t)$ несущих частот в полосе частотной группы лежит в диапазоне (0,7...60), максимальное значение индекса модуляции наблюдается в частотных диапазонах первых трех формант и высокочастотной части спектра РС.

3) При индексах модуляции $\beta(t)$, больших единицы, ширина спектра $S_{\varphi}(\omega, t)$ близка к удвоенной девиации частоты кинем $2\omega k$, на границах равноартикуляционных полос происходит суперпозиция спектров сигналов, что определяет значительную вариативность и неравномерность спектра фонем РС как для шепотной, так и для вокализованной речи.

Результаты проведенных исследований позволяют оценить помехоустойчивость РС, а также адаптировать сигнал маскирования по информационным параметрам кинем, что повышает информационную разведзащищенность выделенных помещений от ТРП при анализе в модели угроз корреляционных методов демаскирования РС.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. ДСТУ 3396.2-97. Державний стандарт України. Захист інформації, Технічний захист інформації. Терміни та визначення. – К., 1998. – 12 с.
2. Фланеган Дж. Анализ, синтез и восприятие речи: Пер. с англ./ Под ред. А. А. Пирогова. – М.: Связь, 1968. – 396 с.
3. Бодуэн де Куртенэ И. А. Разница между фонетикой и психофонетикой. Избранные труды по общему языкознанию, Т. 2, – М., 1963. – 547 с.
4. Вокодерная телефония. Методы и проблемы. / Под ред. А. А. Пирогова – М.: Связь, 1974. – 536 с.
5. Цвикер Э. Ухо как приемник информации. / Э. Цвикер, Р. Фельдкеллер / Пер. с нем. под ред. Б. Г. Белкина. – М.: Связь, 1971. – 225 с.
6. Журавлев В. Н. Анализ гипотезы фазовой чувствительности слухового рецептора / В. Н. Журавлев, А. Е. Архипов // Тези доп. III Міжн. наук. конф. «Інтелектуальні системи прийняття рішень і прикладні аспекти інформаційних технологій (ISDMIT'2007)» – Євпаторія, 2007.
7. Тихонов В. И. Статистическая радиотехника. – М., 1982. – 624 с.
8. Витязев В. В. Вейвлет-анализ временных рядов: Учеб. пособие. – СПб., 2001. – 58 с.
9. Сергиенко А.Б. Цифровая обработка сигналов. – СПб., 2003. – 608 с.
10. Гоноровский И. С. Радиотехнические цепи и сигналы. – М.: Радио и связь, 1986. – 512 с.
11. Михайлов В. Г. Измерение параметров речи / В. Г., Михайлов, Л. В. Златоустова. – М., 1987. – 168 с.

Поступила в редколлегию 30.03.2008.