

УДК 621.055.5

В.Н. Журавлёв, В.С. Кабак

МЕТОДИЧЕСКАЯ И ИНФОРМАЦИОННАЯ АДЕКВАТНОСТЬ ГАРМОНИЧЕСКИХ МЕТОДОВ АНАЛИЗА И ИДЕНТИФИКАЦИИ РЕЧЕВЫХ СИГНАЛОВ

Постановка задачи.

В основе спектральной математической модели речеслухового процесса лежит метод обратимых преобразований речевого сигнала (РС) по системе ортогональных функций. Главное достоинство таких преобразований состоит в том, что они позволяют представить анализируемый сигнал $si(t)$ в виде конечной суммы k независимых детерминированных составляющих $si(t) = \sum_k A_k \psi_k(t)$. При этом, применительно к речеслуховому процессу, синтез и анализ каждой составляющей $A_k \psi_k(t)$ выполняет соответствующий физиологический орган речеслуховой системы. Опуская на данном этапе исследований роль управляющих речеслуховых функций центральной нервной системы человека, проведем анализ адекватности спектральной математической модели идентификации информационных составляющих реализаций речевого сигнала (РС).

Основная часть.

Идея спектральной модели. В существе спектрального подхода можно выделить две основные идеи: синтеза и анализа информационной и идентификационной (по отношению к диктору) составляющих РС. Под спектральным синтезом РС понимается процесс информационного генерирования физиологическими органами речевой системы детерминированных ортогональных составляющих $A_k \psi_k(t)$ с их пространственным суммированием по длине речевого тракта. Количество k ортогональных составляющих и значения их весовых коэффициентов a_k аутентифицируют информационную и идентификационную составляющие РС. Под спектральным анализом, соответственно, понимается процесс информационного разложения физиологическими органами слуховой

системы сигнала $si(t)$ на ортогональные составляющие $si(t) = A_k \psi_k(t)$ с их последующей информационной обработкой в высших отделах ЦНС.

В настоящее время экспериментально установлена и признана [1] неполная информационная адекватность спектрального метода обработки РС. Учитывая требования и ограничения, предъявляемые спектральными методами к параметрам временной функции речевого сигнала $si(t)$, проведем теоретический и экспериментальный анализ методической адекватности спектрального метода идентификации речи.

Наибольшее применение в спектральном анализе получила система комплексных экспоненциальных функций [2]. При известных выражениях базисных функций $\psi_k(t) = \exp(-j\omega t)$ сигнал $si(t)$ однозначно определяется совокупностью коэффициентов A_k и может быть абсолютно точно восстановлен по этим коэффициентам. Таким образом, любая функция гильбертова пространства может быть представлена в виде комплексного ряда Фурье, который лежит в основе спектральной модели РС. При практических применениях ряд Фурье ограничивается определенным количеством членов k . Ограничение числа членов ряда означает аппроксимацию бесконечномерного сигнала k – мерной системой базисных функций спектра сигнала с определенной методической погрешностью в зависимости от фактического спектра сигнала.

Ограничение интервала интегрирования $t \in [Tp]$ временем «квазистационарности» параметров физиологических органов речевой системы приводит к дополнительным (в сумме с вышеуказанной) погрешностям, которые влияют на информационную адекватность. Кроме того, тригонометрические (и связанные с ними комплексно-экспоненциальные) функции являются собственными (детерминированными на интервале интегрирования Tp) функциями и поэтому пригодны для исследования информационных характеристик и параметров сигналов, у которых изменения параметров во времени отсутствуют, т.е. детерминированных.

С позиций анализа РС в частотной области и точного восстановления его после преобразований можно отметить недостатки разложения сигналов в ряды Фурье, отметим основные из них.

Ограниченная информативность анализа на нестационарных временных интервалах и практически полное отсутствие возможностей анализа параметров особенностей сигналов (сингулярностей), т.к. в частотной области происходит «размытие» особенностей сигналов (разрывов, ступенек, пиков и т.п.) как по окну анализа, так и по всему частотному диапазону спектра. Так, например, преобразование Фурье не различает сигнал с суммой двух синусоид (стационарный сигнал), от сигнала с двумя последовательно следующими синусоидами с теми же частотами (нестационарный сигнал), т.к. спектральные коэффициенты a_k вычисляются интегрированием по всему интервалу T_p анализа сигнала, т.е. идентифицируют качественный спектральный состав. Преобразование Фурье в принципе не имеет возможности анализировать частотные характеристики сигнала в произвольные моменты времени.

Детерминированные тригонометрические собственные базисные функции разложения не способны в принципе отображать перепады сигналов с бесконечной крутизной изменения параметра, т.к. для этого требуется бесконечно большое число членов ряда. При ограничении числа членов ряда Фурье в окрестностях скачков и разрывов восстановленного сигнала возникают осцилляции (эффект Гиббса).

Как видно из вышеизложенного, адекватность спектрального метода анализа РС полностью определяется стационарностью сигнала $si(t)$ на анализируемых временных интервалах.

В последующих экспериментальных исследованиях будем определять основные характеристики и параметры стационарности сигнала $si(t, \Delta t)$ [4] как дискретного случайного речевого процесса (СРП), т.е. представим РС в виде числового ряда, отражающего значения функции $si(t)$ через детерминированные дискретные интервалы времени $\Delta t \ll T_p$. Строгое определение стационарности случайного процесса подразумевает неизменность функции распределения плотности вероятностей (ФРПВ) его значений $wp_n[si(t, \Delta t)], t \in (\Delta t), t \in (T_p), \Delta t \in (T_p]$ произвольного порядка n на любых временных интервалах его реализаций $t \in [T_p]$, в частном случае одномерная $wp_1[si(t)]$ ФРПВ

$$wp_1[si(t, \Delta t)] = \frac{\partial F[si(t, \Delta t)]}{\partial si(t, \Delta t)}, \quad (1)$$

где $F[si(t, \Delta t)] = p[si(t, \Delta t) < x]$ - интегральная функция распределения.

Независимость ФРПВ определяет постоянство моментов этих реализаций: среднего и дисперсии, необходимых, но недостаточных условий стационарности. Менее строго (в широком смысле по А.Я. Хинчину [5]) стационарность характеризует неизменность автокорреляционных функций

$$R_{ss}(t, \Delta t, \tau) = \int_0^{Tw} [si(t, \Delta t), si(t, \tau)] d\tau, \tau \in (Tw), \quad (2)$$

где Tw - временной интервал окна анализа реализаций процесса $si(t, \Delta t)$. Корреляционная теория определяет интервал стационарности как время корреляции АКФ τ_0 , который определяется [6] как половина ширины основания прямоугольника единичной высоты, площадь которого равна площади под графиком АКФ:

$$\tau_0 = \frac{1}{2} \int_0^{Tw} R_{ss}(t, \Delta t, \tau) d\tau \Big|_{r_{ss} > 0}. \quad (3)$$

Таким образом, аргументированным требованием, ограничивающим применение спектральных методов как инструмента параметрического информационного анализа, является строгое требование стационарности функции $si(t)$ на исследуемом временном интервале Tr , либо оконных интервалах Tw . Ранжируем порядок исследования требований стационарности оконных реализаций: функция распределения плотности вероятностей, интервал корреляции АКФ среднее и дисперсия РС.

Анализ варибельности ФРПВ (1) однозначно определяет [7] несоответствие требованию стационарности СРП фонем и вывод о неполной информационной адекватности метода спектрального анализа амплитуд РС, однако остается возможность качественного анализа присутствия размытых спектральных составляющих с неопределенными амплитудами. Проведенными исследованиями экспериментально установлен факт несоответствия ФРПВ нормальному закону распределения [8].

Анализ изменения дисперсии РС $R_{ss}(t, \Delta t, \tau)|_{\tau=0}$ и интервала корреляции АКФ (2) показывает их варибельность на времени

длительности фонемы, параметр интервала корреляции (3) чувствителен к верхней граничной частоте спектральной плотности мощности (СПМ) РС [9,10], что подтверждает предыдущий вывод. В результате анализа вышеизложенного можно сделать вывод о несоответствии РС требованиям стационарности на интервале «квазистационарности», что обуславливает:

Неполную информационную и методическую адекватность метода анализа спектра амплитуд. Результаты применения методов спектрального анализа адекватны только в части исследования параметров СПМ РС, при этом необходимо учитывать, что в связи с внутренним методическим анализом АКФ разные фонемы могут иметь одинаковые характеристики и параметры СПМ.

При цифровой обработке РС нельзя применять теорему Котельникова о дискретизации, связанную с верхней частотой анализируемого сигнала.

Для реализации ЦОС нами был предложен и апробирован метод определения частоты дискретизации по параметру доверительной вероятности восстановления формы РС [11], который показал свою эффективность и применялся в последующих экспериментальных исследованиях.

Анализ АКФ показывает её экспоненциально гармонический характер $R_{ss}(t, \Delta t, \tau) = [R_{ss}(t, \Delta t, \tau)|_{\tau=0}]e^{-\tau} \cos \omega_n \tau$, это позволяет сделать вывод об узкополосном энергетическом спектре на несущей частоте ω_n . В связи с этим возникает задача нахождения модулирующей функции кинем $sk(t)$ – сигналов перемещения физиологических органов речевой системы диктора в процессе артикуляции. Информационный сигнал кинем $sk(t)$ является свехнизкочастотным, подтональным ($f_k < 50Hz$) и не может быть передан в атмосфере на жизненно необходимые расстояния, поэтому, в процессе физиологического процесса речеобразования, он модулирует несущие сигналы основного тона (процесс вокализации), обеспечивая необходимую акустическую мощность РС. Представим РС в виде:

$$si(t) = Md_I \left[\sum sk(t), \sum_{i=1}^{g_i} sgi_i(\vec{a}_{m_i}, t) \right], t \in [Tp], \quad (4)$$

где: Md_I – символ системного оператора модуляции источника сообщений, в общем случае нелинейного, \vec{a}_m – вектор параметров m –

го метода модуляции, g_i – количество несущих сигналов. Сигнал фонемы $si(t)$ характеризуется параметрами времени его передачи и полосой частот $\Delta\Omega_i$, определяемой ω_{gi} – частотами несущих сигнала и методом управления информационными параметрами. Зависимости параметра индекса модуляции – Δa_m для амплитудной и частотной модуляции вокализованного РС рассмотрены и исследованы достаточно подробно [12]. На основе гипотезы фазовой чувствительности слуха [13] проведем исследования зависимости изменения параметра Δa_m при угловой φ модуляции несущих сигнала сигналами кинем для вокализованных фонем. Представим сигнал кинем $sk(t)$ как сумму m гармонических сигналов

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

где выражение под знаком косинуса есть

полная фаза, которая является информационным сигналом, т.к. отображает информацию о текущем изменении фазы сигнала кинем. Таким образом, РС $si(t)$, можно представить следующим выражением:

$$si(t) = Md_{\Psi} \left[\sum_{i=1}^m A_i(t) \cos[\omega k_i t + \Psi_i(t)], sgi(\omega_0, t) \right], \quad (5)$$

где $\Psi_i(t)$, ωk_i – фазовый угол и угловая частота i - ой кинемы. Данный сигнал можно анализировать как узкополосный случайный процесс $sw(t) = A(t) \cos[\omega_0 t + \Psi(t)]$ в полосе несущих тональных частот: $\omega_0 \in (\omega^l, \omega^h) \in [\Delta\Omega]$ – нижняя и верхняя частоты анализируемого РС, у которого как огибающая $A(t)$ так и начальная фаза являются случайными функциями, медленно (по сравнению с $\cos \omega_0 t$) изменяющимися во времени. Ранее доказано [12], что РС принадлежат гильбертову пространству с конечной мощностью в интервалах времени Tp и частоты СПМ $N_i(\omega, t), \omega \in (\omega^l, \omega^h) \in (\Delta\Omega), t \in [Tp]$, не содержит дельта функций и разрывов, поэтому для анализа процесса угловой демодуляции РС (5) можно применить метод низкочастотного эквивалента [6]. Угловая демодуляция РС может осуществляться традиционным методом, который предусматривает промежуточное вычисление аналитического

сигнала при помощи преобразования Гильберта и нахождение аргумента $\Psi(\omega k, t)$ исследуемого сигнала.

$$\Psi(\omega k, t) = \arg\{Hil\{A(t) \cos[\omega_0 t + \Psi(t)] \exp(-j\omega_0 t)\}\}, \quad (6)$$

где $Hil\{*\}$ - вычисление преобразования Гильберта, $\arg\{*\}$ - вычисление аргумента функции. Нами учитывалось, что преобразование Гильберта, выполняющее расчет квадратурного дополнения сигнала $si(t)$, оперирует функциями прямого и обратного преобразования Фурье, т.е. требует строгой стационарности сигнала аргумента. При анализе результатов экспериментальных исследований [14], принимая во внимание предыдущие выводы о строгой нестационарности РС, учитывалась их неполная информационную адекватность.

Выводы.

В результате экспериментальных исследований установлен факт разрывности фазовой функции (6) угловой модуляции РС свидетельствует о дискретном характере речеслухового процесса и ставит под сомнение факт полной информационной адекватности метода гармонического анализа и синтеза РС. В дополнение к предыдущему можно констатировать, что современная модель речеслухового процесса не поясняет факта физиологического аналого-цифрового преобразования формы временной реализации РС в низкоскоростную импульсную последовательность сигналов слуховых нейронов ЦНС. Таким образом, основным недостатком рассмотренной гармонической модели РС является её неполная информационная и методическая адекватность эволюционным психофизиологическим речеслуховым процессам, обеспечивающим природный метод кодирования информации в звуковых сигналах речи.

ЛИТЕРАТУРА

1. Галунов В.И. Современные речевые технологии (обзорная статья). [Электронный ресурс] – СПб., 2008. – Режим доступа: <http://www.auditech.ru/article/SpecomRus.doc> свободный. – Загл. с экрана.
2. Журавлев В.Н., Архипов А.Е. Анализ противоречий теорий речеобразования и слуха с позиции идентификации информационных параметров и характеристик речевых сигналов.

- Інформаційні технології та комп'ютерна інженерія № 2(9),
Винница. 2007, С. 180 – 185.
3. Солодовников А.И., Спиваковский А.М. Основы теории и методы спектральной обработки информации. – Л.: Изд-во Ленингр. ун-та. – 1986, 272 с.
 4. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи /Под ред. М.А. Сапожкова. – М. Радио и связь. 1987. – с. 168.
 5. Брандт З. Статистические методы анализа наблюдений. – М.: Мир, 1975, – 312 с.
 6. Френсис Л. Теория сигналов. – М.: Сов. радио, 1974. – 344 с.
 7. Тихонов В.А., Нетребенко К.В. Негауссовы характеристики речевых сигналов. Автоматизированные системы управления и приборы автоматики. Всеукраинский межведомственный научно-технический сборник. Вып. 123. – Харьков: 2003. с. 57 – 62.
 8. Журавлев В.Н., Жуковицкий И.В. Экспериментальные исследования зависимости формы фонем речевого сигнала от их информационного содержания. Міжнародна науково – практична конференція «Сучасні інформаційні технології на транспорті, промисловості та освіті». – Дніпропетровськ, ДНУЖТ: - 2008. - С. 59 - 60.
 9. Журавлев В.Н., Кабак В.С., Рыбин В.О. Анализ параметров автокорреляционной функции фрагментов фонем в задачах идентификации абонента. Радіоелектроніка, інформатика, управління. – 2004. - № 1(9). - С. 29-33.
 10. Кабак В.С., Журавлєв В.Н., Рибин В.О. Підвищення ефективності ідентифікації статичних і динамічних параметрів мовного сигналу. Всеукраїнська науково – практична конференція «Удосконалення діяльності ОВС України з попередження й розкриття злочинів та інших правопорушень». – Запоріжжя, Юридичний інститут МВС України: - 2007. - С. 88 - 89.
 11. Журавлев В.Н. Анализ влияния частоты дискретизации на точность цифровой обработки речевых сигналов в системах биометрической идентификации. Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. – 2005. – № 10. – С. 51-57.
 12. Сорокин В.Н. Теория речеобразования. – М. – Радио и связь. 1985. – с. 312.

13. Журавлев В.Н., Жуковицкий И.В. Анализ методов и результатов исследований зависимости информационного параметра речевого сигнала от фазовых соотношений его гармонических составляющих. Научно – технічний журнал. "Інформаційно – керуючі системи на залізничному транспорті". Харків, - 2007. - №3(65). – С. 8 – 13.
14. Журавлев В.Н., Жуковицкий И.В. Исследование девиации частоты фонации при демодуляционной обработке речевого сигнала. Вестник Днепропетровского национального университета железнодорожного транспорта. Днепропетровск, - 2006.- № 11. - С. 67 - 73.

Получено 18.02.2009г.