

В.Н. Журавлёв, И.В. Жуковицкий

## МОДЕЛЬ ИМПУЛЬСНОГО УПРАВЛЕНИЯ РЕЧЕВЫМ ПРОЦЕССОМ

*Аннотация.* Рассматривается задача, суть которой состоит в раскрытии и анализе импульсного психофизиологического управления речевым процессом, позволяющая повысить достоверность модели речевого источника сообщений.

*Ключевые слова:* речевой сигнал, речеслуховой процесс, информация, энергия, энтропия, синергетика, фонемы, форманты

### Введение

В современных автоматизированных системах большое внимание уделяется удобному интерфейсу операторов АРМов. Замена традиционного ввода информации в АРМ средствами клавиатуры и «мышки» речевым вводом, несомненно, снизит нагрузку оперативного персонала, работающего с АРМами. Поэтому актуальным являются исследования, позволяющие повысить достоверность распознавания речи. Для этих целей необходимо проанализировать как модель источника, так и приемника речевого сигнала.

Основные положения акустической теории речеобразования, предложенной в фундаментальной работе [1] Гельмгольца в 1863 г. остались без изменения до настоящего времени, её развитие изложено в работах Фанта [2] и Соркина [3]. В соответствии с данной теорией, реализация информационной компоненты речевого сигнала осуществляется в процессе управления центральной нервной системой (ЦНС) кинемами артикуляционных физиологических органов, которые регулируют параметры речевого акустического фильтра. Последний, воздействуя на поток воздуха, выходящего из лёгких, синтезирует информационные сигналы – фонемы, которые состоят из суммы гармонических сигналов формант  $s_{for}(\omega, t)$ , определяющих высоту тона речевого сигнала (РС)  $s(\omega, t)$ . В соответствии с гипотезой «фазовой глухоты» речеслуховой системы они определяются как двумерные

случайные сигналы в пространствах изменения параметров их несущей частоты  $\omega_{\text{for}}(t)$  и мощности  $N_{\text{for}}(t)$ :

$$\begin{aligned} s_{\text{for}}(\omega, t) &= f(\omega_{\text{for}}, N_{\text{for}}), t \in [T_p], \\ s(\omega, t) &= \sum_i s_{\text{for}_i}(\omega_i, t) \end{aligned} \quad (1)$$

на информационном временном интервале  $T_p$  длительности звучания фонемы.

Парадигмой информационного анализа сигналов формант РС является исследование функции (1) амплитудного спектра фонем в интервале частот  $\Delta\Omega$  РС:

$$s_{\text{for}}(A, \omega) = \int_t s(\omega, t) \exp(-j\omega t) dt, t \in [T_p], \omega \in [\Delta\Omega]. \quad (2)$$

При этом предполагается, что, в случае принятия гипотезы о локальном постоянстве параметров РС [4], функция распределения плотности вероятностей (ФРПВ) амплитуд РС  $w[s(t)]$  детерминирована и подчиняется нормальному закону распределения. Последнее определяет свойство стационарности сигналов фонем на т.н. «слоговом интервале» – времени  $T_p$  информационной активности фонемы и позволяет применять методы спектрального анализа (2).

В результате экспериментальных исследований ФРПВ амплитуд РС  $w[s(t)]$  установлен факт [5,6] их информационной variability, и несоответствию нормальному закону распределения. Этот факт ставит под сомнение свойство стационарности РС на интервалах времени  $T_p$ , гипотезу о локальном постоянстве параметров речевого тракта и информационную адекватность спектрального метода анализа (2) его амплитуд.

Можно констатировать, что, помимо многих противоречий фактам экспериментальных исследований [7], современная спектральная теория речеслухового процесса не поясняет психофизиологических механизмов синтеза и восприятия высоты тона и громкости РС.

Решение *задачи*, суть которой состоит в *раскрытии и анализе противоречия* дискретного характера информационного психофизиологического управления природным речеслуховым процессом и спектральными методами его анализа, является *целью* исследований настоящей статьи.

### Основная часть

Принято считать, что процессы синтеза и анализа РС осуществляются ЦНС человека дискретно, двоичными импульсными сигналами нейронов. В настоящее время доказано [8], что функционирование эволюционных природных биологических объектов определяется нелинейными стохастическими пространственно-временными диссипативными процессами. Информационное и энергетическое взаимодействие исследуемых процессов целесообразно анализировать, беря в основу *методологию синергетики*, и открытых систем [9] которая позволяет сформулировать следующую *гипотезу* речеслухового дискретного процесса энергоинформационного обмена.

1. Психофизиологические речеслуховые подсистемы диктора и аудитора на интервале времени  $T_p$  речеслухового процесса (РП) синтеза фонемы характеризуются дискретным процессом обмена информацией  $\Delta I(t, \Delta t)$  и энергией  $\Delta E(t, \Delta t)$  между подсистемами в среде функционирования. Эффективная энергия  $E_{\text{eff}}(t, \Delta t) = \int_{T_p} s(t) dt$  (громкость) РС чувствительна к параметрам расстояния  $l_{\text{da}}$  между диктором и аудитором:

$$E_{\text{eff}}(t, \Delta t) = f(l_{\text{da}}), t \in (\Delta t] \in (T_p]. \quad (3)$$

2. Сообщение, представляющее собой информацию  $I(t, \Delta t)$  фонемы РС  $s(t, \Delta t)$ , передается в канал связи дискретными символами энергии  $\Delta E(t, \Delta t) \in s(t, \Delta t)$ . Каждый символ характеризуется своим состоянием – буквой  $\Delta E_k(t, \Delta t_i)$ , которые поступают в канал связи друг за другом в течение интервала времени  $T_p$ . Каждая буква определяется в двумерном непрерывном векторном пространстве энергии и интервала времени передачи символа сообщения  $\Delta E_k(t, \Delta t_i), \Delta E_k \in [\Delta E_{\text{min}}, \Delta E_{\text{max}}], \Delta t_i \in [\Delta t_{\text{min}}, \Delta t_{\text{max}}]$ . Буква переносит соответствующую информацию  $\Delta I_k(t, \Delta t_i) = \Delta I_{k,i}$  фонемы РС, и синтезируется речеслуховой системой диктора методом дискретного управления энергетическими и временными параметрами воздушного потока, нагнетаемого из лёгких.

3. Внутри помещения объекта информационной деятельности (ОИД) действует закон сохранения энергии  $E(t, \Delta t)$ , информации  $I(t, \Delta t)$  и энтропии  $H(t, \Delta t)$ :

$$\begin{aligned} \sum_{k,i} \Delta E_k(t, \Delta t_i) &= \text{const}, \\ \sum_{k,i} \Delta I_{k,i} &= H_0(t) - \sum_{k,i} \Delta H_{k,i}, \end{aligned} \quad (4)$$

где  $H_0(t)$  – начальная энтропия приемника речевых сообщений.

4. Результатом обмена является процесс самоорганизации и уменьшения энтропии  $H_0(t)$  приемника сообщений.

В основе психофизиологических процессов речеслуховых систем источника и приемника сообщений лежит *теория биологических функциональных систем* П.К. Анохина [10]. Методология функционирования открытой биологической функциональной системы речевого источника сообщений в границах объекта исследований предполагают следующую концепцию.

1. Основой речевого процесса (3,4), определяющего энергоинформационный обмен между диссипативными открытыми биологическими информационными системами объекта исследований, являются *системообразующие факторы и функции*.

2. *Системообразующим фактором* психофизиологической речеслуховой системы является факт реализации полезного результата её работы – речеслухового процесса.

3. *Системообразующая функция* – система функциональных или операторных зависимостей энергоинформационных параметров и характеристик речеслухового процесса, обеспечивающих реализацию системообразующего фактора. Параметры системообразующей функции речеслухового процесса и иницируются и формируются в ЦНС посредством синтеза вектора управляющих параметров  $\vec{\mathfrak{X}}_k(t, \Delta t_i)$  сигнала кинемы  $k_k(t, \Delta t_i)$ , который управляет фазовым параметром  $\rho(t, \Delta t_i)$ , влияющим на процесс диссипации энергии  $\Delta E_k(t, \Delta t_k)$  речевого сообщения:

$$f : \Delta I_k(t, \Delta t) \rightarrow \vec{\mathfrak{X}}_k(t, \Delta t_i) \rightarrow k_k(t, \Delta t_i) \rightarrow \rho(t, \Delta t_i) \rightarrow \Delta E_k(t, \Delta t_i). \quad (5)$$

Эффективность изменения вектора управляющих параметров  $\vec{\mathfrak{X}}_k(t, \Delta t_i)$  реализации РП (5) осуществляются под непрерывным контролем ЦНС параметра эффективности  $q_k(t, \Delta t_i)$  и чувствительна к объему апостериорной информации  $I_{k-1}^{\text{ЦНС}}(t, \Delta t_{i-1})$ , содержащейся в накопленном опыте диктора при предыдущих  $(k-1)$ -х реализациях РП:

$$\begin{aligned} q_k(t, \Delta t_i) &= Q[I_k(t, \Delta t_i), I_{k-1}^{\text{ЦНС}}(t, \Delta t_{i-1})], \\ \tilde{\mathfrak{X}}_k(t, \Delta t_i) &= f[q_k(t, \Delta t_i)], \end{aligned} \quad (6)$$

где  $Q[*]$  – системный функциональный оператор анализа информационного расстояния.

Проведем анализ процесса синтеза РС  $s(t, \Delta t)$  с позиции синергетической теории функционирования биологических систем П.К. Анохина.

Источником РС является артикуляционная система диктора. В процессе синтеза РС воздух из легких проходит через трахею, голосовые связки, гортань и затем разветвляется на два потока. Один поток поступает в носоглотку, взаимодействует с носовой полостью и выходит через носовые отверстия. Другой поток, пройдя через ротовую полость, выходит через ротовое отверстие.

Достижение результата в функциональной биологической системе осуществляется с помощью специфических механизмов, из которых наиболее важными являются:

*афферентный синтез* всей поступающей в ЦНС информации;

*принятие решения* с одновременным формированием аппарата прогнозирования результата в виде афферентной модели – акцептора результатов артикуляции;

*собственно действие*;

*сличение, на основе информации обратной связи канала самопрослушивания*, афферентной модели акцептора результатов артикуляции и параметров выполненного речевого процесса;

*коррекция поведения* в случае рассогласования реальных и идеальных (моделированных нервной системой) параметров речевого процесса.

Рассмотрим поэтапный психофизиологический процесс функционирования речевой системы.

*Этап 1.* Афферентный синтез. Начальную стадию поведенческого акта РП составляет афферентный синтез векторов управляющих параметров  $\tilde{\mathfrak{X}}(t, \Delta t_i)$  ЦНС. На этой стадии происходит взаимодействие трех компонент: мотивационного возбуждения необходимости передачи сообщения об информационном образе  $k$ -го объекта  $I_k(O_k)$ , обстановочной афферентации (т.е. информации об интенсивности помех внешней среды  $\xi(t)$ , дальности передачи РС  $l_{\text{да}}(t)$ ) и извлекаемых

из памяти реализаций  $I_{k-1}(O_k)$  прошлого опыта. В результате обработки и синтеза этих компонентов принимается решение об алгоритме управления сигналами кинем  $k_k(t, \Delta t)$  активных артикуляционных органов. Команда, представленная вектором  $\vec{R}_k(t, \Delta t)$ , направляется к артикуляционным исполнительным органам и воплощается в соответствующий РС, состоящий из двух составляющих: информационной и энергетической.

*Этап 2.* Акцептор результатов действия. Необходимой частью функциональной системы является акцептор результатов действия – психофизиологическая обратная связь самоорганизации «артикуляционные исполнительные органы – ЦНС (центр Брока)» – процесс оценки результатов и параметров еще не озвученного РС. Таким образом, еще до осуществления процесса артикуляции РС в ЦНС уже имеется представление о его эффективности, своеобразная модель или образ ожидаемого результата поведения системы.

*Этап 3.* Собственно действие (артикуляция). В процессе артикуляции от центра Брока идут афферентные сигналы (векторов управляющих параметров  $\vec{R}_k(t, \Delta t)$ ) к моторным структурам активных артикуляционных органов, обеспечивающим параметр эффективности (5) реализации системообразующей функции (8) процесса синтеза фонемы. Об эффективности процесса артикуляции сигнализирует поступающая в центр Вернике эфферентная информация  $I_k^{ооо}(O_k)$  от рецепторов слуха по акустическому и костному каналам обратной связи самоорганизации сигналом  $s_k^{ооо}(t, \Delta t)$ , которые регистрируют последовательные этапы выполнения процесса артикуляции (обратная афферентация).

*Этап 4.* Сличение, на основе анализа рецептором слуха сигналов  $s_k^{ооо}(t, \Delta t)$  акустической и костной информационной обратной связи самоорганизации, результатов и параметров артикуляции слова осуществляется в центре Вернике, который хранит базу данных сигналов управления прошлых опытов артикуляции.

*Этап 5.* В случае несоответствия параметра эффективности (6) реализации системообразующей функции (5) процесса синтеза фонемы с определенным ЦНС соответствующим параметром на этапе аф-



ферентного синтеза, производится коррекция векторов управляющих параметров  $\vec{R}_k(t, \Delta t)$  процесса артикуляции в центре Брока.

Приведенный поэтапный анализ процесса речеобразования показывает, что при реализации информационной функции комплексно участвуют три основные системы: речеслуховые зоны ЦНС, речевые физиологические исполнительные органы и органы слуховой системы. Поэтому синергетический анализ процессов речевого обмена необходимо проводить с учетом комплексного взаимодействия указанных систем в единой речеслуховой психофизиологической системе речевого обмена информацией.

На основании вышеизложенного предложим следующую *гипотезу процесса функционирования речевой системы*.

Синергетический энергоинформационный психофизиологический процесс синтеза речевой информации  $I[s(t)]$  акустического сигнала фонемы  $s(t, \Delta t)$  происходит по инициативе и под управлением ЦНС диктора, он состоит из двух, параллельных во времени, взаимообусловленных и неотделимых друг друга, компонент: энергетической  $\Delta E(t, \Delta t) = f^E[s(t, \Delta t)]$  – диссипации и излучения энергии сигнала фонемы и информационной  $\Delta H(t, \Delta t) = f^I[s(t, \Delta t)]$  – снятия части энтропии с источника информации – ЦНС аудитора.

Процесс синтеза сигнала  $k$ -ой фонемы  $s_k(t, \Delta t)$ , переносящего информацию  $I_k[O_k(t)]$ , осуществляется механизмом формирования ЦНС диктора вектора энергетических и информационных управляющих параметров самоорганизации:

$$\vec{\mathfrak{R}}_k(t, \Delta t_i) = f^{E,I}\{I_k[O_k(t)]\}, \quad t \in [\Delta t_i] \quad (7)$$

исполнительных физиологических органов речевой подсистемы. Психофизиологическое действие вектора управляющих параметров заключается в ограничении количества степеней свободы

$$\begin{aligned} nk_k^E(t) &= f^E[\vec{\mathfrak{R}}_k^E(t, \Delta t_i)], \\ nk_k^I(t, \Delta t_i) &= f^I[\vec{\mathfrak{R}}_k^I(t, \Delta t_i)] \end{aligned} \quad (8)$$

исполнительных физиологических органов речевой подсистемы, обеспечивающих энергетические и информационные компоненты процесса синтеза сигнала  $k$ -ой фонемы. В процессе ограничения степеней свободы, исполнительные физиологические органы перестают

совершать хаотические и начинают осуществлять информационные движения самоорганизации, формируя сигналы кинем

$$\begin{aligned} k_k^E(t) &= f^E[nk_k^E(t)], \\ k_k^I(t, \Delta t_i) &= f^I[nk_k^I(t, \Delta t_i)]. \end{aligned} \quad (9)$$

Первые управляют энергетическим параметром избыточного давления воздуха поступающего из лёгких  $\Delta p_k(t)$ . Последние – параметром процесса турбулизации – числом Рейнольдса  $Re(t)$ . При увеличении числа Рейнольдса более критического значения

$$Re_{i,m}(\Delta x_m^I, t, \Delta t_{i,m}, v_{i,m}(t)) > Re^{cr} = f[v_{i,m}(t) \cong \text{const}] \quad (10)$$

ламинарное движение воздуха теряет устойчивость и переходит в качественно другое движение – турбулентное, которое характеризует начало синергетических процессов: увеличения параметра турбулентной вязкости, диссипации кинетической энергии ламинарного потока в энергию ударной волны турбулизации РС.

Процесс диссипации энергии характеризуется импульсным переходом кинетической энергии ламинарного потока в энергию  $\Delta E_{i,m}(t, \Delta t_{i,m})$  сигнала  $k$ -ой фонемы на  $m$ -том информационном участке речевого тракта:

$$s_k(\Delta E_k, t, \Delta t_k) = \sum_m \Delta E_{i,m}(t, \Delta t_{i,m}), \quad t \in [\Delta t_{i,m}] \in [Tp_k]. \quad (11)$$

Анализ (9 – 11) позволяет рассматривать функцию импульсного изменения энергии  $s_k(\Delta E_k, t, \Delta t_k)$  сигнала фонемы как свойство речевой подсистемы реализовать зависимость изменения энтропии сигнала  $k$ -ой фонемы на информационном участке  $\Delta x_m^I$ .

*Системообразующим фактором речевой системы* будет факт реализации функции полного снятия энтропии  $H_k(t, \Delta t_i)$  с источника информации  $I_k(t, \Delta t_i)$  по синтезируемой фонеме  $s_k(t, \Delta t)$  с поддержанием необходимого уровня (громкости) эффективной акустической энергии излучения  $E_{\text{eff}}[s_i(t)] = \text{const}$ :

$$\left\{ \begin{aligned} E_{\text{eff}}[s_i(t)] &= \text{const}, \\ I_i(t, \Delta t_i) \Big|_{t=Tp} &= 0 \end{aligned} \right\}. \quad (12)$$

Системообразующая функция должна отображать последовательность психофизиологического процесса реализации функций (9...14).



Эффективность и точность информационного изменения векторов управляющих параметров  $\tilde{\mathfrak{R}}(t, \Delta t)$  реализации РП осуществляются по бинауральным каналам обратной связи самоорганизации – костным и акустическим. По этим каналам синтезированный РС поступает в подсистему слуха диктора, где, под непрерывным контролем ЦНС, осуществляется определение параметров функции эффективности  $q^{E,I}(t)$  по объему информации  $I_{k-1}^{ЦНС}(t, \Delta t)$  накопленного опыта предыдущих реализаций  $k$ -ой фонемы.

Функциональная схема процесса реализации системообразующей функции приведена на рис. 1.

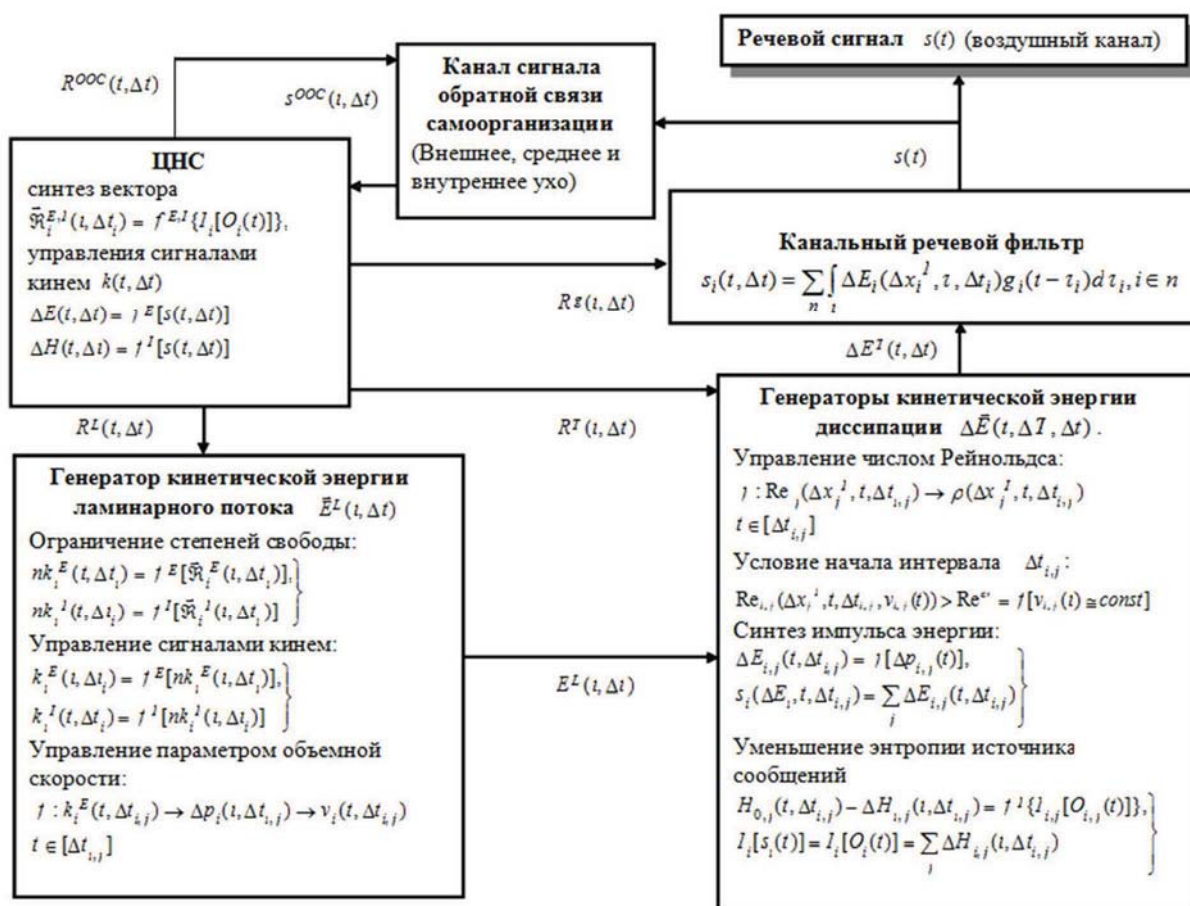


Рисунок 1 - Функциональная схема реализации системообразующей функции речевого процесса

### Выводы

Проведенный теоретический анализ речевого процесса позволяет сделать следующие выводы.

- Аналитически решена задача, заключающаяся в противоречии дискретного характера психофизиологического информационного

управления природным речеслуховым процессом и аналоговыми методами его анализа.

• На базе теории функциональных систем П.К. Анохина разработана импульсная модель психофизиологического синтеза РС. В основе предложенной модели лежат синергетические процессы обмена веществом, энергией и информацией, которые базируются на эффекте дискретного изменения динамической вязкости воздуха при фазовом переходе его течения в управляемых ЦНС сечениях речевого тракта.

• Предлагаемая модель является расширением и объединением существующих моделей речеобразования на основе методологий синергетики и биологических функциональных систем. Современные модели речеобразования и слуха, базирующиеся на методе интегральной обработки параметров и характеристик процесса речевого обмена и основанные на ограничениях гипотезы о локальном постоянстве параметров речевого тракта, представляют частный случай предлагаемой модели.

#### ЛИТЕРАТУРА

1. Helmholtz H. von, Die Lehe von Tonempfindungen. Brannschweig, Vieweg, 1863.
2. Сорокин В.Н. Теория речеобразования. – М. – Радио и связь. 1985. – 312 с.
3. Фант Г. Акустическая теория речеобразования: Пер. с англ./Под ред. В.С. Григорьева. – М.: Наука, 1964. – 284 с.
4. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. – М. Радио и связь. 1985. – 176 с.
5. 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.
6. 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.
7. 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.
8. Николис Г., Пригожин И. Самоорганизация в неравновесных системах. – М.: Мир. – 1979. – 512 с.
9. Климонтович Ю.Л. Статистическая теория открытых систем. – М: ТОО «Янус». – 1995. – 624 с.
10. Анохин П.К. Кибернетика функциональных систем // Избранные труды. – М: Медицина. – 1998. – 400 с.