

ИСПОЛЬЗОВАНИЕ МЕТОДА КОМПЕНСАЦИИ ИНФОРМАЦИОННЫХ ПОТОКОВ ПРИ ПОСТРОЕНИИ СИСТЕМ УПРАВЛЕНИЯ С РЕЧЕВЫМ ИНТЕРФЕЙСОМ

Фанина Л.А., Бражник Д.А.

Введение.

Несмотря на почти полувековую историю исследований в области речевых технологий, задача распознавания речи до сих пор остается открытой. На сегодняшний день существует множество программных продуктов в той или иной мере реализующих системы распознавания речи, однако качество их достаточно сомнительное.

Анализ существующих проблем и перспективы их решения.

Почти все современные реализации систем распознавания речи имеют один очень важный недостаток – *зависимость от диктора*.

Существенной проблемой является *подавление стационарных и нестационарных помех*. В настоящее время системы диктовки текстов и голосового управления практически не применяют в своем функционировании алгоритмы шумоподавления. Это связано с тем, что компьютерные речевые системы разрабатываются для работы в офисе или дома, где уровень внешних помех не очень высок. Однако в системах речевого управления реальными техническими устройствами, например в авиации, используют шумоподавление. Так как отсутствие шумоподавления в компьютерных речевых системах влияет на надежность распознавания (например, глубокий вдох пользователя система идентифицирует как одно из слов словаря), можно ожидать в ближайшем будущем внедрения алгоритмов и устройств шумоподавления в персональные компьютеры.

Проблема перехода к распознаванию непрерывной речи связана в основном с недостатком вычислительных мощностей персональных компьютеров, делающим системы непрерывной диктовки слишком дорогими и потому непопулярными. Ожидается повсеместный переход на распознавание непрерывной речи в ближайшее десятилетие, причем задача распознавания дискретной речи не потеряет своей актуальности.

На сегодняшний день *проблема учета контекста* (синтаксиса, семантики) при восстановлении последовательности произнесенных слов пользуются простыми, не зависящими от языка грамматиками, позволяющими в достаточной степени учитывать связь слов в предложении. В будущем ожидается усложнение грамматик, разработанных с учетом специфики языков, и разработки соответствующих процедур обучения.

Одной из существенных проблем является *поиск новых звуковых параметров*. В настоящее время для распознавания речи в основном используют спектральные параметры речи – быстрое преобразование Фурье, спектр линейного предсказания, кепстральные коэффициенты и т.п. Эти параметры обладают как рядом преимуществ (соответствие восприятию звука человеком, возможность применения Евклидовой метрики или ее вероятностного аналога – расстояния Махаланобиса для сравнения окон), так и рядом недостатков (зависимость спектральных параметров от голоса диктора). Продолжаются исследования по поиску инвариантных относительно типа голоса и влияния шумов речевых параметров.

Проблемой *поиска новых алгоритмов* восстановления последовательности произнесенных звуков занимаются многие исследовательские лаборатории. На сегодняшний день для сравнения последовательности акустических параметров с эталонами слов словаря используется практически три метода – самый распространенный метод скрытых марковских моделей, тесно связанный с ним метод динамической деформации времени (применяемый на словарях относительно малого размера) и стоящий несколько поодаль метод с использованием нейронных сетей. В связи с бурным развитием речевых техноло-

гий ожидается появление принципиально новых математических методов в области распознавания речи.

Проблема выделения слов в потоке слитной речи при распознавании речевых сигналов. Решение данной задачи предполагает ее анализ, результатом чего является определение характеристик речевого сигнала (на уровне фонем, слов), т.е. параметры модели линейного предсказания, динамика изменения этих параметров, частота основного тона, анализ частотных областей и т.д. При этом возникает проблема выделения фрагментов речи в потоке слитной речи.

Следовательно, к современным системам распознавания речи можно выдвинуть следующие требования:

- высокое качество распознавания,
- достаточно высокое быстродействие,
- устойчивость к внешним шумам и вариации параметров диктора,
- минимальные затраты времени на обучение и настройку системы распознавания.

Цель данной статьи – создание алгоритма обеспечивающего инвариантность системы по отношению к внешним возмущениям.

Теоретические исследования.

Для образа $\omega \in \Omega$ и координаты x рассмотрим меру близости $\rho(\omega, x)$, в качестве которой используем нормированную взаимную корреляционную функцию между сигналом образа и эталона [1].

$$\rho(\omega, \omega^*) = 1 - R(f_\omega(t), f_{\omega^*}(t)) \quad (1)$$

Исходя из метода допустимых преобразований [1], учтем управления эталоном

$$\rho(\omega, \omega^*, u) = 1 - R(f_\omega(t), f_{\omega^*}(t, u)) \quad (2)$$

Учитывая, что метод допустимых преобразований предполагает соответствие эталона и объекта, введем ограничение $\omega \sim \omega^*$, данное ограничение ведет к тривиальному равенству $f_\omega(t) = f_{\omega^*}(t)$. С другой стороны, учитывая, что в случае выполнения ограничения, взаимная условная информация, получаемая системой равна нулю [2], можем перейти к ограничению вида

$$I(\omega / \omega^*) = 0 \quad (3)$$

Что позволяет рассматривать задачу

$$\begin{aligned} \omega^*, u^* &\rightarrow \min \{1 - R(f_\omega(t), f_{\omega^*}(t, u))\}; \\ I(\omega / \omega^*) &= 0. \end{aligned} \quad (4)$$

Учитывая, что для близких образов, как функция цели, так и ограничения выпуклы, получаем задачу выпуклого программирования с функцией Лагранжа

$$L(\omega, u, \lambda) = 1 - R(f_\omega(t), f_{\omega^*}(t, u)) + \lambda I(\omega / \omega^*). \quad (5)$$

Множитель Лагранжа в данной задаче определяет чувствительность корреляционной функции к выбору эталона

$$\lambda(u) = \frac{\partial(1 - R(f_{\omega}(t), f_{\omega^*}(t, u)))}{\partial I(\omega/\omega^*)} = - \frac{\partial(R(f_{\omega}(t), f_{\omega^*}(t, u)))}{\partial I(\omega/\omega^*)}. \quad (6)$$

Поскольку в данной задаче множитель Лагранжа естественно связан с управлением, двойственной переменной является управление эталоном.

Таким образом, получаем прямую и двойственную задачи в виде:

$$\begin{aligned} u^* &\longrightarrow \min_{\omega=\omega^*} L(\omega, u, \lambda); \\ \omega^* &\longrightarrow \max_{u=u^*} L(\omega, u, \lambda). \end{aligned} \quad (7)$$

Учитывая структуру функции Лагранжа в данной задаче можно записать

$$\begin{aligned} u^* &\longrightarrow \sup_{\omega=\omega^*} R(\omega, \omega^*(u)) = 0; \\ \omega^* &\longrightarrow \inf_{u=u^*} I(\omega, \omega^*(u^*)) = 0. \end{aligned} \quad (8)$$

Прямая задача порождает корреляционные методы распознавания, но по своей структуре прямая задача наилучшим образом подходит к решению задачи позиционирования сигнала образа в окне контроля.

Естественный интерес представляет анализ возможности построения системы с использованием корреляционных методов, включая задачу выделения области единичного образа и задачу принятия решения о классификации сигнала.

Двойственная задача лежит в основе метода компенсации информационных потоков [3].

Принцип компенсации информационных потоков исходит из достаточно простых предположений:

- во-первых, система компенсирует внешние возмущения,
- во-вторых, если система может воспроизвести все то, что происходит во внешней среде, то возможно построение системы распознавания работающей без ошибок.

Критерием качества распознавания является то, что система точно отображает внешние воздействия, т.е. $I(X, Y) = 0$, причем при $I(X, Y) \rightarrow \min \Rightarrow R(X, Y) \rightarrow \max$.

При построении систем распознавания в качестве целевого функционала можно использовать вероятность правильного распознавания. Оценка вероятности ошибки связана с оценкой информационного потока соотношением [1]:

$$H(K/M) \leq H(K/D) \leq H_{\max}(H(K/M)/2) \quad (9)$$

где $H(K/M)$ - энтропия по описанию при K классах и M описаниях, мера недостаточности описания M по отношению к идеальному описанию, безошибочно указывающему класс; $H(K/D)$ - энтропия по оптимальному решению, где оптимальное решение можно рассматривать как некоторое предельно упрощенное описание, указывающее к какому классу с наибольшей вероятностью принадлежит данный сигнал; $H_{\max}(H(K/M)/2)$ - максимальная энтропия я при P_{\max} .

В силу связи точной верхней и нижней границ энтропий [2] максимальная вероятность правильного распознавания достигается при $H(K/M) = H(K/D) = 0$. Следовательно, рассматривая входной сигнал как систему X , а эталон как его описание Y , можно утверждать, что условная энтропия $H(X/Y)$ обращается в ноль, если эталон совпадает с входным распознаваемым сигналом, т.е. отсутствует неопределенность эталона по отношению к сигналу. Таким образом, в рамках задачи распознавания, принятие решения по минимуму условной информации с учетом связности объекта и эталона, доставляет минимум вероятности ошибки.

Рассматривая систему с эталонной моделью Ω_0 и конкретным образом речевого сигнала Ω_i , выделим основные потоки информации в системе, движение которой происходит в пространстве состояния x под управлением u . Информационный поток источника и информационный поток эталонной модели порождаются физически различными источниками, но они связаны единством образов Ω^* и Ω_i . Их условная информация, таким образом, связана и обладает экстремальными свойствами, т.е. существуют такие x^* и u^* , которые доставляют минимум условной информации I_{ω, ω^*} .

Рассматривая потоки I_0 и I_i как встречные с мерой отклонения I_ε , где $I_\varepsilon = I_\varepsilon(x, u, \Omega_i, t)$ - условная информация I_{ω, ω^*} . В данной постановке решение задачи оптимального управления для эталона $\Omega_i: x^*, u^* \rightarrow \min I_\varepsilon$ связано с решением задачи распознавания в классе образов $\Omega: \Omega_i \rightarrow \inf I_\varepsilon$.

Одним из существенных моментов в методе компенсации информационных потоков является оценка взаимной информации по затратам управления в задаче допустимых преобразований.

Количество информации является отражением одного объекта другим и мерой соответствия состояний этих объектов.

Рассмотрим множество допустимых управлений U_0 и соотнесем этому множеству объем допустимых управлений V_0 . В таком случае принятие решения u_i на i -том шаге уменьшает объем управлений до V_i , следовательно, отношение V_i/V_0 при геометрической интерпретации вероятности определяет вероятность принятия ошибочного решения q . Тогда вероятность принятия правильного решения определяется как

$$p_i = 1 - q_i = 1 - \frac{V_i}{V_0}. \quad (10)$$

С другой стороны количество информации служащее основанием для принятия решения при заданном эталоне можно определить как

$$I_{\omega, \omega^*} = -\ln(p_i) = -\ln\left(1 - \frac{V_i}{V_0}\right). \quad (11)$$

Представим логарифм в виде ряда

$$I_{\omega, \omega^*} = -\ln\left(1 - \frac{V_i}{V_0}\right) = \frac{V_i}{V_0} + \frac{V_i^2}{2V_0^2} + \dots + \frac{V_i^n}{nV_0^n} + \dots. \quad (12)$$

Информация, получаемая за m шагов определится, как сумма информации полученной на каждом шаге

$$\tilde{I}_{\omega, \omega^*} = \frac{1}{V_0} \sum_{j=1}^m \sum_{i=1}^n \frac{V_j^i}{i} = \tilde{u}. \quad (13)$$

Что определяет полные приведенные затраты управления в задаче принятия решения об управлении эталоном. Подтверждение целесообразности данного подхода обосновано в работе [2].

Тогда двойственная задача может быть сформулирована в виде, где можно непосредственно оценивать взаимную информацию

$$\omega^* \xrightarrow[u=u^*]{} \inf \tilde{u}(\omega, \omega^*(u^*)). \quad (14)$$

Для проверки теоретических предположений проведено моделирование как прямой, так и двойственной задачи.

Экспериментальная часть.

Для анализа возможности построения системы с использованием двойственности задачи рассмотрим ансамбль достаточно близких речевых сигналов, рисунок 1.

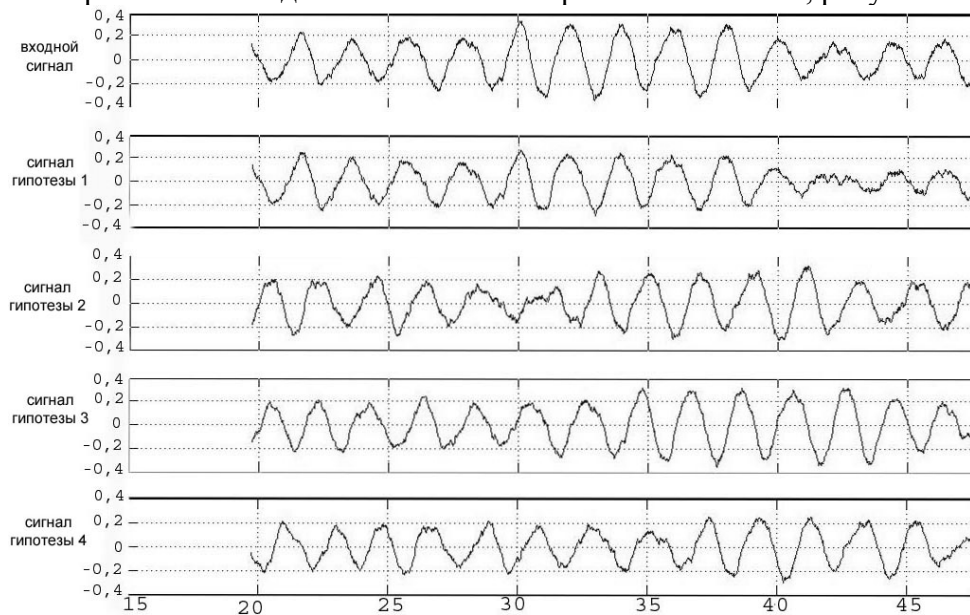


Рис. 1 Реализации тестовых сигналов

Для оценки корреляционных свойств ансамбля определим взаимные корреляции, на основе модели в среде MATLAB, рисунок 2.

Полученные оценки корреляции указывают на возможность использования данного метода, но одновременно подчеркивают типичность корреляционной функции речевого сигнала – колебательность, что определяет целесообразность использования фазовых методов привязки. Однако, даже для упрощенной модели необходимо центрировать сигналы и выполнять длинные операции умножения, что усложняет использование метода.

С другой стороны, двойственную задачу можно рассматривать как компенсацию информационных потоков на входе системы входного акустического сигнала (набор векторов его параметров) и эталона хранящегося в базе знаний системы распознавания речи (априорная информация).

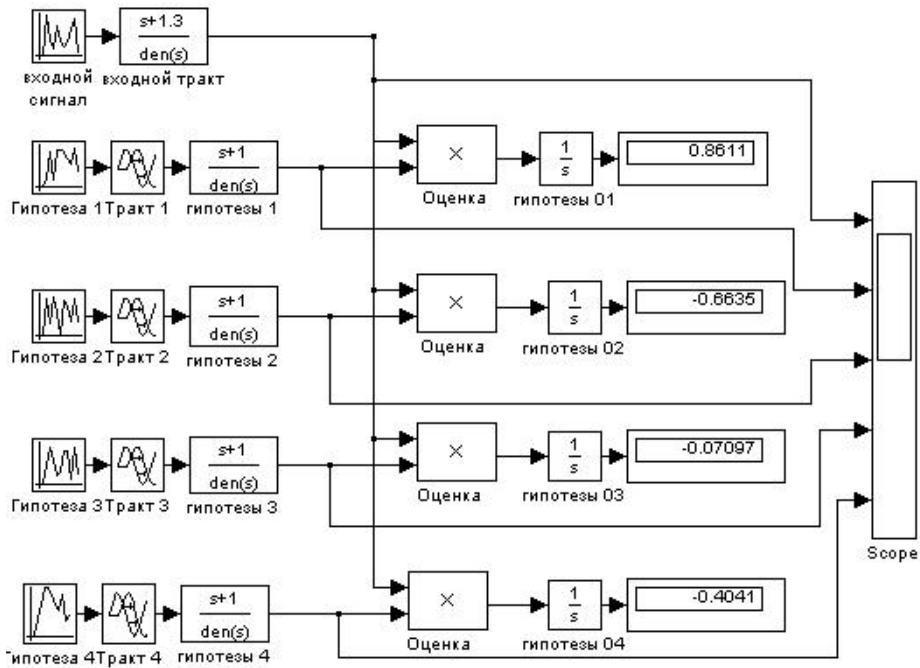


Рис. 2 Схема для оценки корреляционных свойств ансамбля тестовых сигналов

Учитывая, что при минимуме взаимной информации затраты на управление эталонным минимальны, используем в качестве меры близости оценку затрат управления в задаче воспроизведения сигнала эталона при ограничении определяемом входным сигналом системы. Такая постановка позволяет использовать типовой дельта-модулятор как основной элемент системы. На рисунке 3 приведена схема моделирования работы элемента сравнения.

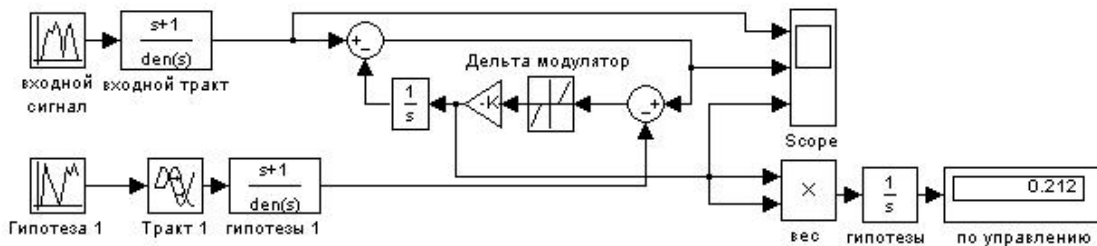


Рис. 3 Схема моделирования элемента сравнения

На рисунке 4 приведены результаты моделирования.

Приведенный на рис. 4а входной сигнал повторяется системой практически без искажений рис 4б, затраты на управление незначительны рис. 4с, так как выбран эталон близкий входному сигналу.

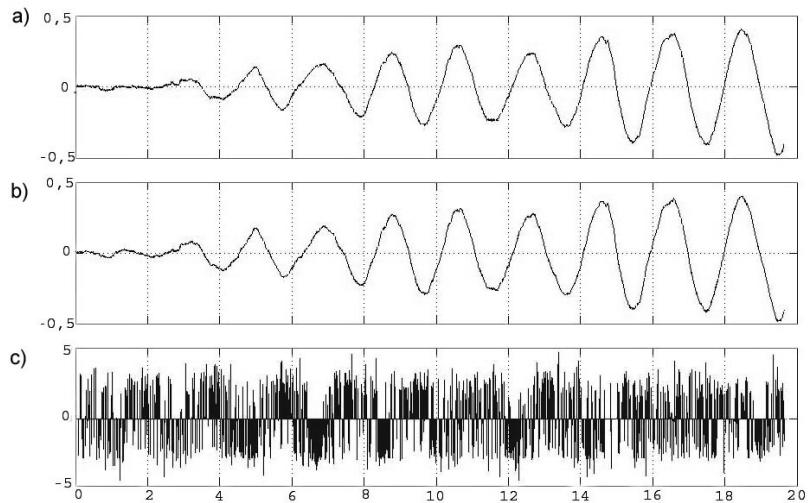


Рис. 4. Результаты моделирования элемента сравнения.

Однако при обработке сигналов далеких от эталона затраты управления резко возрастают. Для оценки возможностей метода проведено моделирование системы на ансамбле сигналов, использованном ранее для оценки корреляционного метода. Схема моделирования построена с выделенными элементами сравнения для каждого сигнала гипотезы, рисунок 5.

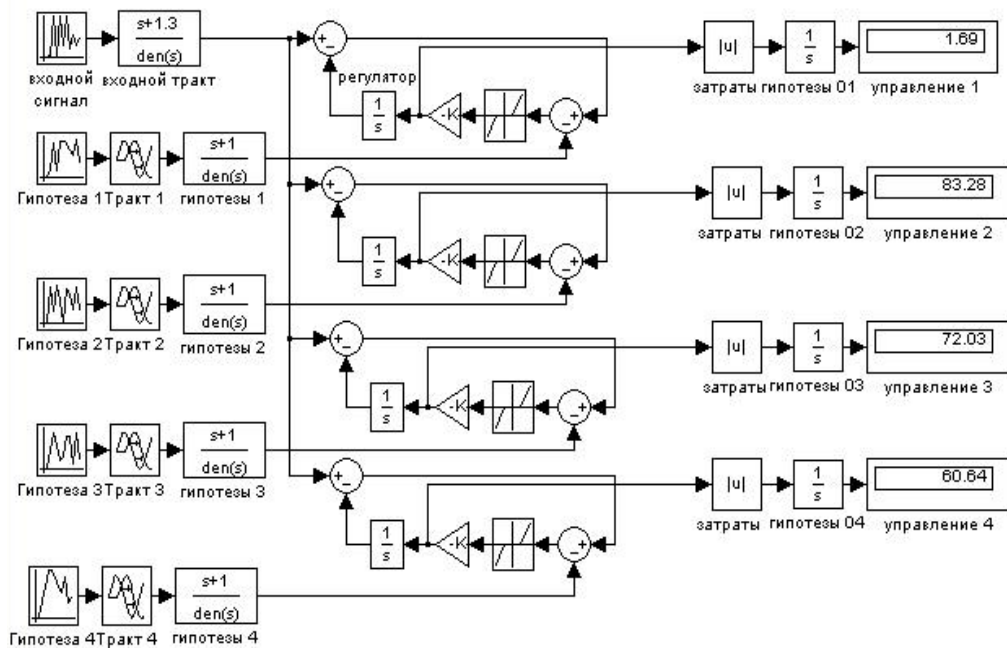


Рис. 5 Схема моделирования затрат по управлению

Как видно из результатов моделирования использование прямой задачи позволяет реализовать принцип компенсации информационных потоков с использованием простых средств. Полученные значения оценок гипотез по управлению разнесены гораздо больше, по сравнению с корреляционным методом.

Таблица 1

№ гипотезы	1	2	3	4
Значение корреляционной функции	0,8611	-0,6653	-0,07091	-0,4041
Затраты на управление	1,69	83,28	72,03	60,64

Результаты моделирования позволяют обосновать структуру системы распознавания речи как систему с компенсацией информационных потоков.

Учитывая, что в принципе компенсации необходима реализация встречного потока информации, в качестве генератора встречного потока используем систему синтеза речи. Тогда структурная схема системы принимает простой вид



Рис. 6 Обобщенная структурная схема системы на основе принципа компенсации

Естественным недостатком данного подхода является сложность формирования первых гипотез. Данный недостаток свойственен всем системам, построенным по отклонению, и вызывает значительные трудности, связанные с необходимостью учета времени переходных процессов системе. Таким образом, для обеспечения формирования первичных гипотез целесообразно ввести в систему дополнительный канал, построенный по принципу возмущения – любая достаточно работоспособная система распознавания речи. В этом случае структура принимает вид известный в теории управления как адаптивная инвариантная система.

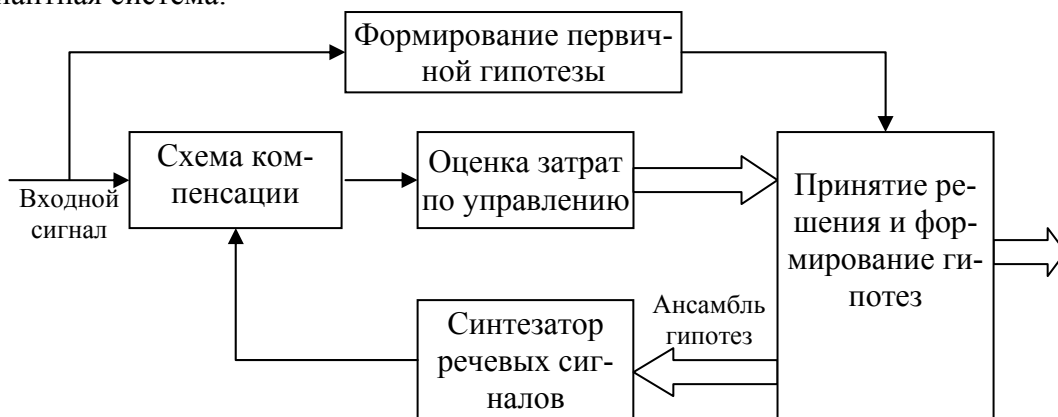


Рис. 7 Обобщенная структурная схема системы на основе принципа компенсации с формированием первичной гипотезы

Окончательный ответ распознавания основывается на результате сравнения этих двух информационных потоков. На рис. 8 представлена концептуальная схема распознавания речи на основе принципа компенсации информационных потоков.

Блок логики формирования гипотез подает ансамбль наиболее вероятных гипотез (глубина гипотез ограничена разрядной сеткой) на блок генерации сигналов, которые затем поступают через дельта-модулятор на схему компенсации. Отклонение ε через отрицательную обратную связь также поступает на вход дельта-модулятора. В качестве критерия используются затраты на управление эталоном для компенсации внешнего сигнала.

При формировании и сравнении описаний речевых сигналов возникает необходимость их нормирования. Применяются методы нелинейного нормирования изложенные в [4]. Выделение границ слова основано на рассмотрении вводимых параметров в некоторое “окно” длиной в несколько отсчетов. Выделение начала и конца произнесения производится программно с использованием буфера, который позволяет исключить ложные срабатывания системы от коротких шумовых помех, а также потери части слова из-за смы-

чек. С этой целью в алгоритм введены логические правила, позволяющие отбросить реализацию, если она содержит количество отсчетов меньше заданного, и закончить ввод, если длительность паузы после окончания реализации превышает заданную.

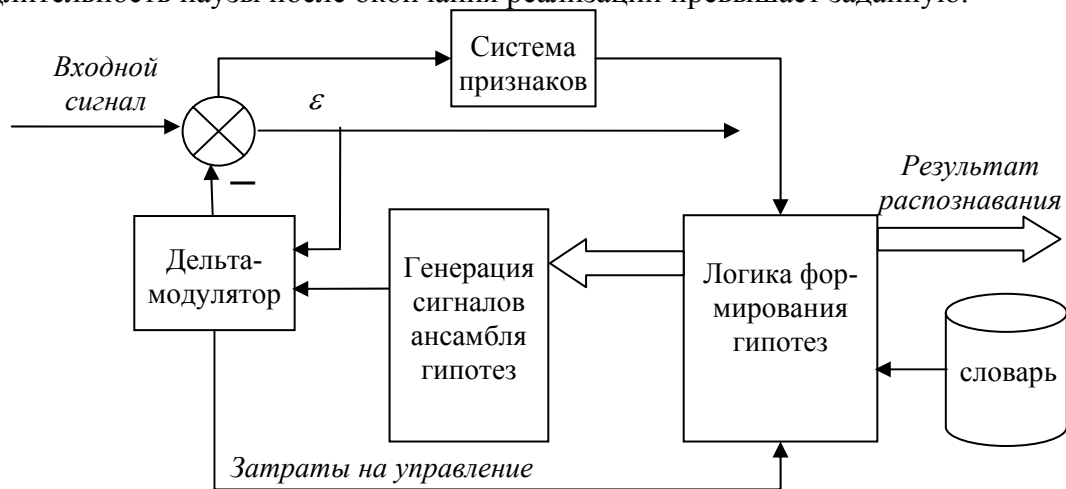


Рис. 8 Концептуальная схема распознавания речи на основе принципа компенсации информационных потоков

Следует подчеркнуть, что предложенная структура является общей концепцией для всех методов распознавания. Методы построения элементов системы не будут влиять на ее структурное решение.

Выводы:

1. Построение системы распознавания речевых сигналов можно рассматривать как прямую задачу минимизации взаимной информации, либо как двойственную задачу максимизации взаимной корреляции.
2. Оценить взаимную информацию между эталоном и входным сигналом возможно по затратам управления на решение задачи восстановления сигнала на основе ресурса эталона.
3. Инвариантность системы распознавания возможно обеспечить при построении системы проверки гипотез с использованием метода компенсации информационных потоков для оценки противоречивости гипотез.
4. Результаты моделирования позволяют считать, что даже на ансамблях с высокой взаимной корреляцией использование метода компенсации информационных потоков более предпочтительно.
5. Рассматриваемая структура является системой построенной по отклонению, что позволяет компенсировать возмущения без анализа их природы.

Article is devoted to the decision a problem of speech recognition. The direct and dual problems are considered. The results of theoretical and experimental research are presented.

1. Методы оптимальных решений в распознавании изображений, В.А. Ковалевский. «Наука», -М., 1976
2. Рожков С.А. Исследование и разработка системы фильтрации сигналов дефектов ткани с выделением признаков их классификации: Диссертация на соискание ученой степени кандидата технических наук: 05.13.07.- Херсон. 1997. –223 с.
3. Храпливый А.П., Рожков С.А., Терновая Т.И., Бражник Д.А. Использование принципа компенсации информационных потоков при создании систем распознавания
4. Плотников В.П. Речевой диалог в системах управления. – М.: Машиностроение, 1988.