

Система оперативной модификации голоса диктора на основе полувокодера

А.С. Рылов,

доктор технических наук

В.В. Киселёв,

директор ООО «Речевые технологии»

А.Г. Давыдов,

кандидат технических наук, научный сотрудник

В.А. Чижденко,

старший научный сотрудник



В работе рассматривается вопрос модификации голоса диктора на основе полувокодера. Предлагается два варианта системы модификации голоса диктора, позволяющие выполнять нелинейное изменение характеристик голоса диктора таким образом, что последующая идентификация инструментальными средствами становится невозможной.

Abstract

The paper deals with voice modification on the basis of a semi-vocoder. We suggest two variants of a voice modification system which provide non-linear modification of the voice characteristics, so that subsequent identification using instrumental means becomes impossible.

1. Введение

Модификация голоса диктора является специфической задачей, востребованной, например, системами синтеза речи по тексту. В подобных системах процедура модификации применяется для обеспечения большего диапазона настройки голоса диктора — если необходимо из подготовленного голоса одного диктора получить голос другого диктора. В англоязычной литературе данная процедура называется voice morphing.

Другой областью применения систем модификации голоса диктора (СМГ) является защита голоса свидетеля при даче показаний в суде. В этом случае модификация голоса диктора должна проводиться в реальном времени с обеспечением минимальной задержки преобразования и невозможности обратного преобразования.

Таким образом, СМГ могут использоваться в самых различных областях, при этом к ним могут предъявляться различные требования. Классификация СМГ по их назначению представлена на *рисунке 1*.

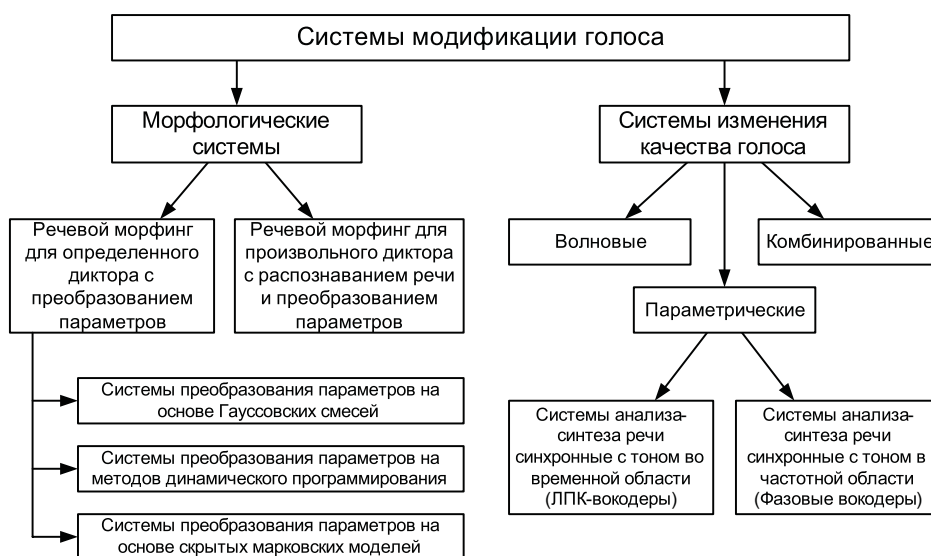


Рис. 1. Классификация систем модификации голоса диктора

Как видно из приведенного рисунка, все СМГ можно разбить на два типа: морфологические и системы для изменения качества голоса.

К первому типу относятся системы, которые позволяют трансформировать голос исходного диктора в голос определённого диктора (диктора-«мишени»). При этом осуществляется адаптация параметров речи исходного диктора к параметрам речи диктора-«мишени» для определённого речевого фрагмента (предложения, фразы), который был произнесён обоими дикторами. К наиболее сложным морфологическим системам относятся системы с распознаванием речи произвольного диктора.

Неморфологические СМГ предназначены для того, чтобы изменять качество голоса диктора, не добиваясь внешней схожести на голос какого-либо определённого диктора. По своему устройству эти СМГ можно разделить на три категории: волновые, параметрические и комбинированные, т.е. такие, в которых используются манипуляции с формой речевой волны и какими-либо акустическими параметрами речи. К параметрическим системам относятся системы анализа-синтеза речи (вокодеры), работающие во временной либо частотной областях, в которых осуществляется преобразование параметров, отвечающих за формантную структуру сигнала и за просодию речи.

Методы достижения поставленной цели — модификации голоса диктора — для морфологических и неморфологических СМГ существенно различаются. Это связано с тем, что исходные условия решения задачи являются различными. При модификации голоса диктора для систем синтеза речи по тексту доступным является большой объём подготовленных записей исходного голоса диктора (записанных с большим соотношением сигнал/шум, размеченных на фонемы и периоды основного тона). Запись голоса целевого диктора, как правило, также является качественной. Время решения задачи преобразования голоса одного диктора в голос другого диктора (время вычислений) для систем синтеза речи по тексту не является критическим и может равняться нескольким суткам. Дополнительным требованием, предъявляемым при решении такой задачи, является высокое качество полученного результата, а задача использования необратимого преобразования исходного голоса диктора в целевой голос диктора не ставится.

При модификации голоса диктора в оперативных условиях голос исходного диктора, как правило, является неизвестным, однако и голос целевого диктора обозначается в значительной степени условно. Например, требуется понизить тембр голоса или повысить тембр голоса. Дополнительным требованием в этом случае является невозможность обратимого преобразования для получения голоса исходного диктора. Вне зависимости от области применения системы модификации голоса диктора, обязательной задачей является сохранение высокой разборчивости произнесённого текста.

Данная статья посвящена рассмотрению неморфологической системы необратимой высококачественной модификации голоса диктора в оперативных условиях.

2. Система модификации голоса диктора на основе искажения линейных спектральных частот

Структурная схема системы модификации голоса диктора на основе искажения линейных спектральных частот (ЛСЧ, LSF в литературе на английском языке [1]) приведена на [рисунке 2](#). На рисунке 2 использованы следующие обозначения:

x, x' — последовательность отсчётов входного и выходного сигналов;

F_x — частота дискретизации входного и выходного сигналов;

F_s — частота дискретизации сигнала, на которой работает блок анализа;

$[s]$ — кадр сигнала;

ν — мера тона для каждого кадра сигнала;

$\vec{a} = \{a_i\}, \vec{a}' = \{a'_i\}$ — исходный и модифицированный векторы коэффициентов линейного предсказания, где $i = \overline{1, M}$, M — порядок модели предсказания;

$[e], [e']$ — исходный и модифицированный кадры ошибки предсказания;

T_e, T'_e — исходная и модифицированная длительность кадра ошибки предсказания;

$\vec{\omega} = \{\omega_1\}, \vec{\omega}' = \{\omega'_1\}$ — исходный и модифицированный векторы ЛСЧ.

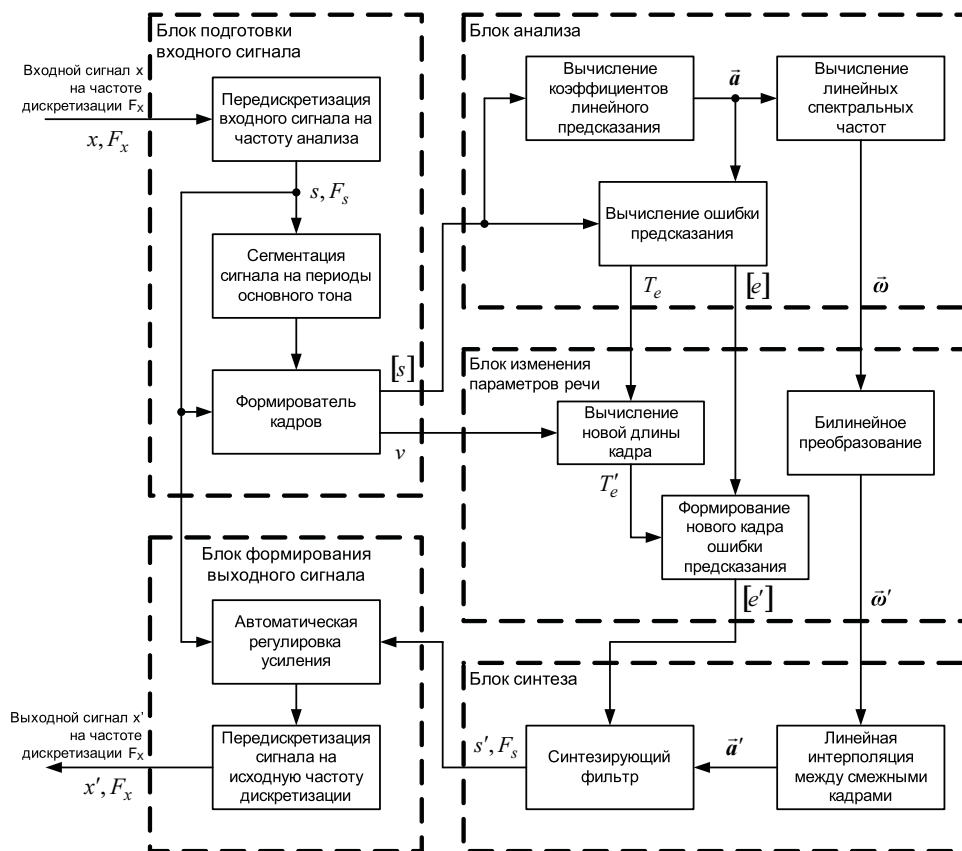


Рис. 2. Структурная схема системы модификации голоса диктора на основе искажения линейных спектральных частот

Как видно из структурной схемы, система включает блок предварительной подготовки входного сигнала, блок анализа (вычисления параметров речевого сигнала), блок модификации параметров, блок синтеза (формирования изменённого речевого сигнала), блок формирования выходного сигнала.

2.1. Блок подготовки входного сигнала предназначен для передискретизации входного сигнала x на частоту дискретизации F_s , формирования кадров $[s]$ речевого сигнала (РС) и вычисления меры тона ν (степени вокализованности) каждого кадра.

Оптимальная частота анализа, определённая в ходе экспериментальных исследований, оказалась близка к 11025 Гц.

Формирование кадров анализа $[s]$ выполняется таким образом, что на вокализованных участках речи каждый кадр содержит один период основного тона. На невокализованных участках речи сигнал разделяется на кадры размером 20 мс.

Параллельно с каждым формируемым кадром сигнала $[s]$ формируется признак вокализованности кадра — ν . Данная операция выполняется в соответствии

с алгоритмом, рассмотренным в [2] и модифицированным для применения в условиях реального масштаба времени.

2.2. Блок анализа служит для вычисления вектора коэффициентов линейного предсказания речи \vec{a} [3, 4] для каждого кадра РС [s]. В соответствии с найденными коэффициентами линейного предсказания вычисляется вектор ЛСЧ \vec{w} [1]. Вычисление кадра ошибки предсказания [e] для найденных коэффициентов линейного предсказания \vec{a} выполняется фильтром с конечной импульсной характеристикой, линия задержки которого не сбрасывается при переходе от одного кадра анализа к другому. Кадр ошибки предсказания [e], его длительность T_e и вектор ЛСЧ \vec{w} передаются в блок изменения параметров речи.

2.3. Блок изменения параметров речи предназначен для формирования модифицированных кадра ошибки предсказания [e'] и вектора ЛСЧ \vec{w}' .

2.3.1. Изменение кадра ошибки предсказания для вокализованных и невокализованных кадров речевого сигнала выполняется по немного различным формулам. Для вокализованных кадров новая длительность кадра T_e равна:

$$\hat{T}_e = T_e \cdot k_{T_0mul} \cdot k_{\Delta T_0mul} + \bar{T}_e \cdot k_{T_0mul} \cdot (1 - k_{\Delta T_0mul}),$$

где k_{T_0mul} — коэффициент изменения длительности периода основного тона; $k_{\Delta T_0mul}$ — коэффициент изменения производной длительности периода основного тона; \bar{T}_e — среднее значение длительности периода основного тона диктора, уточняемое в процессе анализа. Коэффициенты k_{T_0mul} и $k_{\Delta T_0mul}$ задаются оператором.

Вычисление длительности нового кадра ошибки предсказания для невокализованного кадра выполняется в соответствии со следующей формулой:

$$\hat{T}_e = T_e \cdot k_{last_mul},$$

где k_{last_mul} — коэффициент изменения длительности последнего вокализованного кадра.

Значение k_{last_mul} при запуске программы устанавливается равным k_{T_0mul} , а в процессе работы на каждом вокализованном кадре постоянно обновляется в соответствии с формулой $k_{last_mul} = \hat{T}_e / T_e$. Это позволяет маскировать ошибки классификации вокализованных кадров как невокализованных (что иногда наблюдается при малом отношении сигнал/шум на краях вокализованных участков РС).

Для более надёжного маскирования истинного поведения кривой периода основного тона длительность как вокализованных, так и невокализованных кадров ошибки предсказания может (при установке соответствующих настроек оператором) изменяться в соответствии со следующими формулами:

$$\hat{F}_0 = \frac{1}{\hat{T}_e}; F'_0 = \begin{cases} \hat{F}_0 \cdot \left[1 + \frac{k_{T_0vm}}{100} \cdot \sin\left(\frac{\hat{F}_0}{10} + k_{T_0vf} \cdot 2\pi t\right) \right], & \text{если } \hat{F}_0 < 100; \\ \hat{F}_0 + k_{T_0vm} \cdot \sin\left(\frac{\hat{F}_0}{10} + k_{T_0vf} \cdot 2\pi t\right), & \text{если } \hat{F}_0 \geq 100; \end{cases} T'_e = \frac{1}{F'_0}.$$



Как видно из приведённых формул, данный вид маскирования кривой частоты основного тона напоминает её частотную модуляцию, а задаваемые оператором коэффициенты k_{T_0Vm} и k_{T_0Vf} определяют величину девиации частоты и частоту модулирующего сигнала.

Значение длительности кадра ошибки предсказания T'_e дополнительно ограничивается таким образом, чтобы получившееся значение не отличалось более чем в два раза от исходного значения длительности кадра ошибки предсказания T_e .

Формирование нового кадра ошибки предсказания в соответствии с найденным новым значением длительности выполняется: при увеличении длины кадра — линейной интерполяцией, при уменьшении длины кадра — передискретизацией.

Пример изменения частоты основного тона для задания различных значений управляющих коэффициентов приведён на [рисунке 3](#). Значение частоты основного тона оценивалось сторонним программным средством [5].

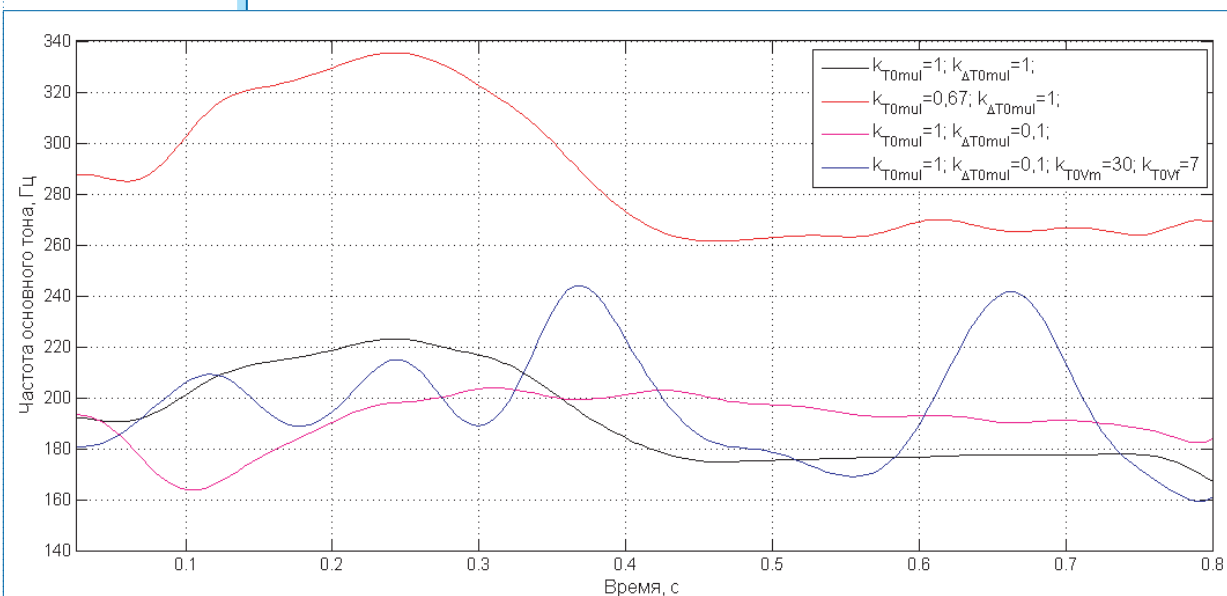


Рис. 3. Модификация частоты основного тона для различных управляющих коэффициентов

2.3.2. Выбор способа модификации огибающей спектра РС в значительной степени определяет качественные характеристики всей системы искажения голоса диктора. Очевидно, что наибольшую свободу в модификации голоса диктора можно получить при работе с фонограммами речи. При работе в реальном времени большое значение играет допустимая величина задержки между подачей исходного голоса и формированием изменённого. Чем эта задержка больше, тем более сложные алгоритмы можно использовать для модификации голоса, однако при этом система становится всё менее пригодной для использования в режиме диалога. Для модификации огибающей спектра РС в процессе разработки системы были исследованы три способа:

- I. Модификация огибающей спектра РС посредством искажения индексов разностей ЛСЧ на основе билинейного преобразования [6].
- II. Искажение коэффициентов ЛСЧ на основе билинейного преобразования.
- III. Искажение спектра посредством изменения угловых координат полюсов.

Краткое описание первых двух способов модификации спектра и полученных результатов, а также анализ достоинств и недостатков приводится в данном подпункте описания системы модификации голоса диктора. Более подробное рассмотрение способа модификации огибающей спектра РС посредством изменения угловых координат полюсов приводится в следующем разделе.

- I. Алгоритм модификации огибающей спектра РС посредством искажения индексов разностей ЛСЧ включает следующие этапы:

1. Вычисление вектора коэффициентов ЛСЧ текущего кадра РС $\vec{\omega} = \{\omega_i\}$ и вектора

коэффициентов ЛСЧ, равномерно расположенных по частоте $\vec{\omega} = \{\omega_i\} = \frac{i}{M+1} \pi$

(что соответствует сигналу с равномерным распределением спектральной плотности мощности).

2. Вычисление разности $\Delta \vec{\omega} = \vec{\omega} - \vec{\omega} \equiv \{\Delta \omega_i\} = \{\omega_i\} - \frac{i}{M+1} \pi$.

3. Искажение индексов вектора $\Delta \vec{\omega}$ как своеобразного представления шкалы частот по формуле

$$i' = \frac{M-1}{\pi} \cdot \left[\phi_\alpha \left(\frac{i-1}{M-1} \pi \right) + 1 \right], \text{ где } \phi_\alpha(\omega) = \omega - 2 \cdot \arctg \left[\frac{\alpha \cdot \sin \omega}{1 + \alpha \cdot \cos \omega} \right] \text{ — билинейное}$$

преобразование, α — управляющий коэффициент преобразования, задаваемый оператором и определяющий вид и степень модификации огибающей спектра РС. Пример функции билинейного преобразования для нескольких значений управляющих коэффициентов приведён на [рисунке 4а](#).

4. Вычисление значений последовательности $\Delta \omega_i$ с абсциссами i' в точках i посредством квадратичной интерполяции (получение модифицированного вектора разностей $\Delta \vec{\omega}' = \{\Delta \omega'_i\}$).

5. Вычисление модифицированного вектора ЛСЧ как суммы модифицированного вектора

разностей $\Delta \vec{\omega}'$ и вектора $\vec{\omega}$: $\vec{\omega}' = \Delta \vec{\omega}' + \vec{\omega} \equiv \{\omega'_i\} = \{\Delta \omega'_i\} + \frac{i}{M+1} \pi$.

Пример искажения огибающей спектра РС в соответствии с этим алгоритмом приведён на [рисунке 4б](#).

При практической реализации обнаружилось, что рассмотренный алгоритм должен быть дополнен этапом проверки и коррекции значений модифицированного вектора ЛСЧ $\vec{\omega}'$. Это связано с тем, что в результирующем векторе иногда наблюдается значительное сближение двух смежных коэффициентов (что в огибающей спектра РС отражается как резкое увеличение мощности на частоте этих коэффициентов). Особенно это характерно для коэффициентов, отвечающих за первую форманту. Также при использовании рассмотренного выше алгоритма иногда получается вектор $\vec{\omega}'$, компоненты которого не следуют по возрастанию. Чаще всего наблюдается случай, когда первый коэффициент оказывается больше второго, а остальные следуют по возрастанию. Это соответствует нестабильному фильтру, полюс которого вышел

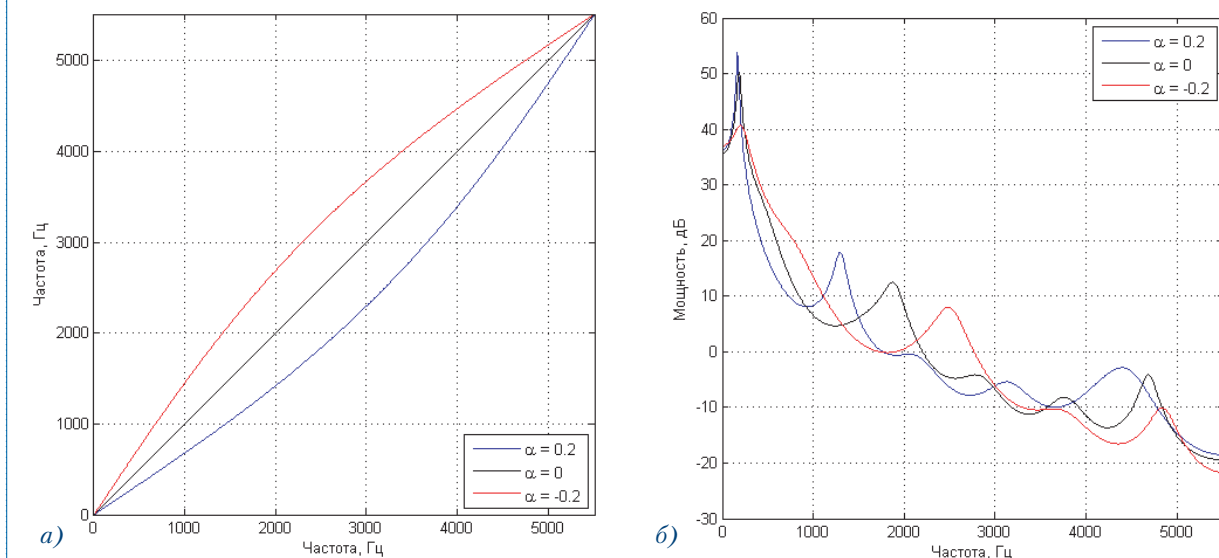


Рис. 4. Функция билинейного преобразования (а) и пример искажения огибающей спектра РС (б) для трёх значений управляющего коэффициента α

за единичную окружность. Для коррекции этих особенностей алгоритм целесообразно дополнить следующим этапом.

6. Все смежные пары коэффициентов вектора $\vec{\omega}'$, расстояние между которыми

отвечает условию $\omega'_{i+1} - \omega'_i < \omega_{\Delta \min}$, заменяются на $\omega''_i = \frac{\omega'_i + \omega'_{i+1}}{2} - \frac{\omega_{\Delta \min}}{2}$

и $\omega''_{i+1} = \frac{\omega'_i + \omega'_{i+1}}{2} + \frac{\omega_{\Delta \min}}{2}$. Для тестовых записей,

сделанных с частотой дискретизации 11025 Гц, и порядка модели предсказания $M = 14$ минимально допустимое расстояние $\omega_{\Delta \min}$ было оценено как приблизительно равное 0,05.

II. Второй способ модификации голоса диктора заключается в искажении непосредственно коэффициентов ЛСЧ по формуле билинейного преобразования $\phi_\alpha(\omega)$. Этот способ, в отличие от предыдущего, не требует дополнительного этапа проверки и коррекции. Пример результата искажения огибающей спектра РС данным способом приведен на [рисунке 5](#).

Как видно из приведённого рисунка, при таком способе модификации подвергается не только расположение максимумов огибающей спектра, но и наклон спектра.

2.4. Блок синтеза служит для выполнения линейной интерполяции векторов ЛСЧ между смежными кадрами, которая используется для сглаживания пере-

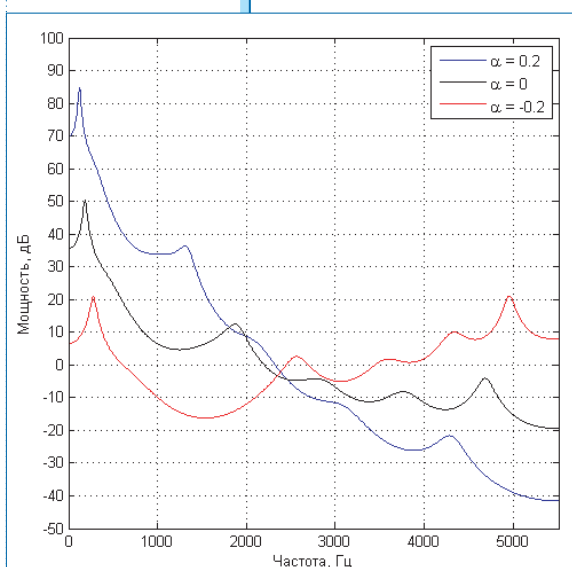


Рис. 5. Модификация огибающей спектра РС путём билинейного преобразования коэффициентов ЛСЧ

ходов на границах синтезируемых кадров. При вычислении каждого отсчёта синтезируемого сигнала путём линейной интерполяции определяется текущее значение вектора ЛСЧ, которое преобразуется в коэффициенты синтезирующего фильтра.

Синтезирующий фильтр, в отличие от анализирующего, применяемого при вычислении кадра ошибки предсказания, является фильтром с бесконечной импульсной характеристикой, который возбуждается модифицированным вектором ошибки предсказания $[e']$.

2.5. Блок формирования выходного сигнала служит для автоматической регулировки усиления синтезированного сигнала (необходимой при втором способе модификации огибающей спектра РС) и его передискретизации на частоту дискретизации входного сигнала.

3. Модификация голоса диктора на основе искажения полюсов передаточной характеристики

Искажение огибающей спектра РС при помощи билинейного преобразования не позволяет выполнять независимую модификацию некоторых формант без искажения остальных формант. Преодолеть это ограничение можно при использовании модификации голоса диктора на основе искажения угловых координат полюсов передаточной характеристики.

Отличие системы модификации голоса диктора на основе искажения полюсов передаточной характеристики от схемы, приведённой на рисунке 2, заключается в блоке изменения параметров речи, в котором вместо билинейного преобразования используется искажение угловых координат полюсов в соответствии со структурной схемой, приведённой на [рисунке 6](#).

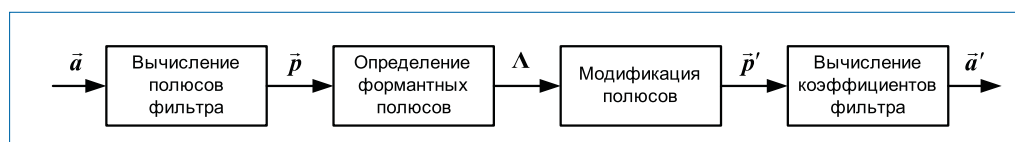


Рис. 6. Схема модификации полюсов передаточной характеристики

Как видно из данного рисунка, вектор коэффициентов предсказания \vec{a} преобразуется в вектор полюсов $\vec{p} = \{p_k\}$, $k = \overline{1, K}$, из которых рассматриваются только те, которые находятся в верхней полуплоскости. На множестве найденных полюсов p_k определяются те, которые определяют частоты и амплитуды формант РС (определение формантных полюсов). Результатом этой процедуры является матрица соответствия формант и полюсов Λ . В соответствии с матрицей Λ угловые координаты полюсов подвергаются модификации для получения нового вектора полюсов $\vec{p}' = \{p'_k\}$, который затем преобразуется в новый вектор коэффициентов фильтра \vec{a}' . Процедуры вычисления полюсов фильтра по его коэффициентам и коэффициентов по полюсам достаточно подробно рассмотрены в литературе [7].

Процедура определения формантных полюсов включает следующие этапы:

1. Вычисление матрицы расстояний D между полюсами p_k и полюсами p_{Fn} ($n = \overline{1, N}$, $N = 5$), имеющими единичную амплитуду и угловые координаты, соответствующие середине частотных диапазонов формант. Для удобства дальнейшего рассмотрения будем считать,



что полюса p_k отсортированы в порядке возрастания их аргумента: $arg(p_{k-1}) < arg(p_k) < arg(p_{k+1})$.

Частотные диапазоны изменения формант хорошо изучены и могут быть найдены, например, в [8, 9]. При разработке СМГ были использованы частотные диапазоны формант, приведённые в таблице 1.

Таблица 1

Частотные диапазоны формант

Номер форманты	F_1	F_2	F_3	F_4	F_5
Частотный диапазон, Гц	100–1000	500–3000	1500–4000	2500–5300	3500–5500

Как видно из таблицы 1, частотные диапазоны формант были несколько увеличены (по сравнению с данными, приводимыми в [8, 9]) для исключения ошибок. Частотный диапазон пятой форманты был ограничен 5500 Гц в связи с использованием для анализа РС с частотой дискретизации, равной 11025 Гц. Таким образом, матрица расстояний D вычисляется в соответствии со следующей формулой:

$$D = [d_{nk}] = [|p_{Fn} - p_k| \cdot (1 - |p_k|)].$$

Для полюсов, лежащих за границами частотных диапазонов соответствующей форманты, расстояние d_{nk} принималось равным бесконечности. Результирующая матрица расстояний D может принять, например, следующий вид:

$$D = \begin{pmatrix} d_{11} & \infty & \infty & \infty & \infty & \infty \\ d_{21} & d_{22} & d_{23} & \infty & \infty & \infty \\ \infty & \infty & d_{33} & \infty & \infty & \infty \\ \infty & \infty & d_{43} & d_{44} & d_{45} & \infty \\ \infty & \infty & \infty & d_{54} & d_{55} & d_{56} \end{pmatrix}.$$

С учётом вышеизложенного, задача определения формантных полюсов сводится к выбору такого множества Λ элементов матрицы расстояний D , которое удовлетворяет следующим условиям.

1. Из каждой строки n матрицы расстояний D в конечное множество выбирается один элемент d_{nk} .
2. Так как полюса p_k отсортированы в порядке возрастания их аргумента, из каждой строки матрицы расстояний D может быть выбран только тот элемент d_{nk} , индекс столбца k которого лежит в диапазоне $[n; n + K - N]$.
3. Индексы столбцов выбранных элементов должны быть отсортированы в порядке возрастания: для любых двух различных элементов d_{nk} и d_{ml} , если $n < m$, то $k < l$; обратное также верно.

4. Сумма элементов d_{nk} множества Λ должна быть минимальной: $\rho = \sum_{n=1}^N d_{nk} \rightarrow \min$.

Перечисленные условия являются достаточно жёсткими и, как правило, оставляют очень мало различных вариантов для определения формантных полюсов, вследствие чего даже применение алгоритма полного перебора является вычислительно приемлемым.

Для модификации голоса диктора используются полюса вектора $\vec{p}_{\Lambda} = \{p_{\lambda}\}$, индексы λ которых совпадают со вторыми индексами элементов d_{nk} множества Λ .

Использование приведённых выше условий позволяет приблизительно в 90% случаев правильно определить формантные полюса. Однако на переходных участках РС, при малом отношении сигнал/шум, встречаются случаи, когда формантные полюса определяются неверно. Признаком таких случаев является резкое изменение угловой координаты формантного полюса от предыдущего кадра к текущему. Расстояние между формантными полюсами Θ предыдущего кадра \vec{p}_{Λ}^{t-1} и текущего кадра \vec{p}_{Λ}^t определяется так:

$$\Theta(\vec{p}_{\Lambda}^{t-1}, \vec{p}_{\Lambda}^t) = \sum_{\lambda=1}^N \ln(|p_{\lambda}^t - p_{\lambda}^{t-1}| + 1).$$

Если расстояние оказывается больше порогового значения (в ходе экспериментов определённого как приблизительно равное единице), то путём последовательного удаления из рассмотрения полюсов p_{Fb} , $b = 2, \bar{N}$ выбирается тот вариант определения формантных полюсов, при котором достигается минимальное значение суммы ρ .

Для модификации формантных полюсов можно использовать самые различные алгоритмы, однако даже модификация голоса диктора путём задания смещений угловых координат формантных полюсов позволяет получить вполне удовлетворительные результаты.

4. Заключение

Разработанная система модификации голоса диктора позволяет оперативно решать поставленную задачу таким образом, что восстановление исходного голоса диктора становится невозможным. Это достигается за счёт использования нелинейного изменения длительности периода основного тона, которое управляется коэффициентами, задающими среднюю длительность, дисперсию и модуляцию периода основного тона. Таким образом, голос диктора может быть с различной степенью преобразован в низкий либо высокий, монотонный либо эмоциональный. Для дополнительной защиты от возможности восстановления исходного голоса диктора коэффициенты могут задаваться как функции времени.

Проведённые предварительные эксперименты показали, что даже при значительной модификации система обеспечивает высокое качество и разборчивость модифицированного голоса.

Билинейное преобразование линейных спектральных частот, использованное в рассмотренной системе для модификации формантных частот, показывает хорошие результаты искажения голоса.

Модификация голоса диктора на основе искажения полюсов передаточной характеристики хотя и является более сложной для реализации, чем билинейное преобразование, но обеспечивает большие возможности искажения голоса.



Литература

1. Huang X., Acero A., Hon H.-W. Spoken Language Processing: A Guide to theory, algorithm, and system development. — New Jersey: Prentice Hall. — 2001. — 1008 p.
2. Лобанов Б.М., Давыдов А.Г. Алгоритм высокоточной разметки на питчи элементов компиляции для синтеза речи по тексту // Компьютерная лингвистика и интеллектуальные технологии (Диалог 2007): труды Международной конференции, Бекасово, 30 мая-3 июня 2007г. / Институт проблем информатики РАН; редкол.: Л.Л. Иомдин [и др.]. М.: Издательский центр РГГУ, 2007. С.388-392.
3. Маркел Дж.Д., Грэй А.Х. Линейное предсказание речи: Пер. с англ./Под ред. Ю.Н. Прохорова, В.С. Звезда. М.: Связь, 1980. 308 с., ил.
4. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения: Пер. с англ. М.: Мир, 1990. 584 с., ил.
5. Camacho A., Harris J.G. A sawtooth waveform inspired pitch estimator for speech and music // The Journal of the Acoustical Society of America. — 2008. — Vol. 124, issue 3. — P. 1638–1652.
6. G. Strecha, O. Jokisch, M. Eichner, R. Hoffmann Codec Integrated Voice Conversion for Embedded Speech Synthesis // Interspeech 2005, Lisbon, Portugal, September 4–8, 2005. P. 2589–2592.
7. Самарский А.А., Гулин А.В. Численные методы: Учебн. пособие для вузов. М.: Наука. Гл. ред. физ.-мат. лит., 1989. 432 с.
8. Фант Г. Акустическая теория речеобразования: Пер. с англ. Л.А. Варшавского и В.И. Медведева / Под ред. В.С. Григорьева. М.: Наука, 1964. 284 с., ил.
9. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи / Под. ред. М.А. Сапожкова. М.: Радио и связь, 1987. 168 с., ил.

Рылов Александр Александрович —

доктор технических наук, Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники».

Киселёв Виталий Владимирович —

директор ООО «Речевые технологии», г. Минск. С 1999 г. профессионально занимается системами синтеза и распознавания речи, диалоговыми речевыми системами. Автор более 25 научных публикаций в области речевых технологий. Основные научные интересы связаны с системами обработкой и анализом текста и речи, системами синтеза, распознавания речи, поиска ключевых слов.

Давыдов Андрей Геннадьевич —

кандидат технических наук, старший научный сотрудник Академии управления при Президенте Республики Беларусь. Основные научные интересы связаны с областью цифровой обработки сигналов, анализом, сжатием и распознаванием речи. Автор более 20 научных публикаций и 3 патентов.

Чижденко Виктор Анатольевич —

старший научный сотрудник ООО «Речевые технологии», г. Минск. Основные научные интересы связаны с областью цифровой обработки сигналов, анализом, сжатием и распознаванием речи. Автор более 10 научных публикаций и 1 патента.

Обращаем Ваше внимание на факт безвременной кончины Рылова Александра Александровича.