Конверсия голоса с использованием модели сепарации речевого сигнала на компоненты «гармоники+шум» и переходные фреймы

А.Н. Павловец,
аспирант
М.З. Ливщиц,
кандидат технических наук, доцент
Д.С. Лихачёв,
кандидат технических наук, доцент
А.А. Петровский,
доктор технических наук, профессор

В статье представлена система конверсии голоса, основанная на модели сепарации речевого сигнала на «гармоники+шум» и переходные фреймы с раздельной конверсией для каждой компоненты модели. Преимущество системы конверсии голоса в данном случае складывается из достоинств анализа-синтеза гармонической модели с достоинствами анализа и конверсии переходных сегментов во временной области. Неформальные тесты прослушивания показали, что узнаваемость диктора соответствует приблизительно 70%, реконструированная речь характеризуется достаточно высокой разборчивостью.

Abctract
The voice conversion system based on the harmonic+noise speech signal model is presented in the given paper. One of the critical tasks in voice conversion framework is speaker parameter estimation. Here the method based on the Harmonic-Noise-Transient (HNT) decomposition of speech is proposed with the idea of processing each of the components separately and further converting them separately.

1. Введение

Проблема конверсии голоса становится очень популярной в мире. Сущность конверсии заключается в модификации голоса диктора, являющегося в данном случае источником, в голос другого (целевого) диктора. Актуальность темы исследований обусловлена широким применением устройств конверсии голоса в мультимедиа-системах реального времени: синтез речи по тексту (устранение «компьютерного акцента»); виртуальное дублирование (восстановление звуковых дорожек кинофильмов); защита свидетелей (применение в судебной практике); оперативная смена диктора в коммуникационных системах (озвучивание SMS-сообщений в мобильных телефонах) [1].

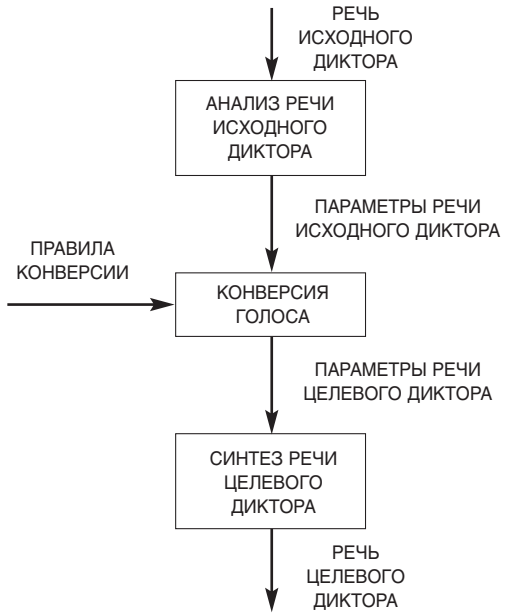

Рис. 1. Типовая схема процесса конверсии голоса

Общий алгоритм процесса конверсии показан на ***рис. 1***.

Процесс конверсии голоса можно разбить на два этапа: обучения и конверсии. На первом из них (этапе обучения) выделяется множество характеристических параметров исходного и целевого дикторов и определяются правила конверсии, посредством которых параметры исходного диктора будут преобразовываться в параметры целевого диктора. На втором этапе (этапе конверсии) характеристики речи исходного диктора преобразуются с использованием правил, определённых на первом этапе.

При реализации системы конверсии голоса требуется решить два основных вопроса: как и какие параметры извлекать из речевого сигнала, подлежащего преобразованию, и как модифицировать эти параметры таким образом, чтобы преобразованная речь была похожа на речь целевого диктора. В работе [2] улучшен подход [3], который был основан на модели «гармоники+шум», путём использования декомпозиции анализируемой речи на вокализованную и шумоподобную компоненты. Ранее подобный метод был применён в области кодирования речи [4, 5]. Дальнейшие исследования показали, что модель [2] не является достаточной в полной мере, поскольку с её помощью нельзя корректно анализировать переходные сегменты речи. Это послужило причиной дополнения модели [2] режимом анализа переходных сегментов.

Определение вокализованных областей в речевом сигнале является довольно сложной задачей. Вокализованность может определяться для анализируемого сегмента речи в целом [6], также можно находить максимальную частоту вокализованности [3] или принимать решение по вокализованности в какой-либо полосе частот [7]. Проблема заключается в том, что принимаемое решение обычно имеет два варианта: область либо вокализована, либо нет. С точки зрения процесса речеобразования более точным было бы рассматривать вокализованную речь как сумму вокализованной и шумоподобной составляющих. В [8] был предложен метод декомпозиции на вокализованную и шумоподобную компоненты, который применяется к сигналу —

остатку линейного предсказания. Идея заключается в использовании итеративного алгоритма, основанного на последовательном применении ДПФ/ОДПФ для определения шумовой компоненты.

Ещё один метод декомпозиции заключается в использовании гармонического фильтра, параметры которого изменяются в зависимости от частоты основного тона [9]: речевой сигнал взвешивается окном, при этом в окно должно укладываться некоторое целое количество периодов основного тона. Так же, как в работах [8] и [9], в подходе, представляемом в данной статье, считается, что вокализованная и шумоподобная составляющие присутствуют во всём диапазоне частот. Спектральный анализ проводится в области гармоник фундаментальной частоты речи, для этого ДПФ было модифицировано таким образом, чтобы учитывать изменение её контура. Точность определения параметров модели, а именно частоты основного тона, амплитуд и фаз гармоник, повышена за счёт использования метода анализа через синтез. Предполагается, что гармоническая компонента определяется суммой гармоник фундаментальной частоты с изменяющимися во времени амплитудами и фазами. Декомпозиция выполняется во временной области, шумоподобная компонента определяется разностью между оригинальной речью и синтезированной гармонической компонентой.

Для модификации параметров речи было предложено большое количество статистических подходов. Популярность приобрели методы, основанные на векторном квантовании. В этом случае определение правил конверсии представляет собой установление соответствия между кодовыми книгами, представляющими акустические классы дикторов. В [10] был описан метод, основанный на жёсткой кластеризации и дискретном соответствии между кодовыми книгами. Получаемый характеристический вектор y'_t в момент времени t определяется путём квантования исходного характеристического вектора и подстановки вместо него соответствующей центроиды из кодовой книги целевого диктора.

Однако жёсткая кластеризация подразумевает большую ошибку квантования. В данной работе представлена модификация метода [10]. В зависимости от типа сегмента речи используются различные кодовые книги. Целью работы является обеспечение лучшего качества конверсии голоса с использованием модели сепарации речевого сигнала на компоненты «гармоники+шум» и переходные фреймы.

2. Модель сепарации речевого сигнала на компоненты «гармоники+шум» и переходные фреймы

2.1. Гармонический анализ

В предлагаемой модели речевой сигнал представляется так:

$$s(i) = h(i) + r(i) + t(i)\,, \qquad (1)$$

где $h(i)$ — вокализованная (гармоническая) компонента, $r(i)$ — сигнал-остаток вокализованной компоненты (шум), $t(i)$ — переходный фрейм. Режим работы анализатора речевого сигнала полностью определяется наличием либо отсутствием основного тона (***рис. 2***).

Данная модель была успешно апробирована в области кодирования речи и аудиосигнала, например, [11, 12]. Первая из упомянутых работ представляет гибридный кодер речи, который сочетает параметрический кодер, работающий в частотной области (для

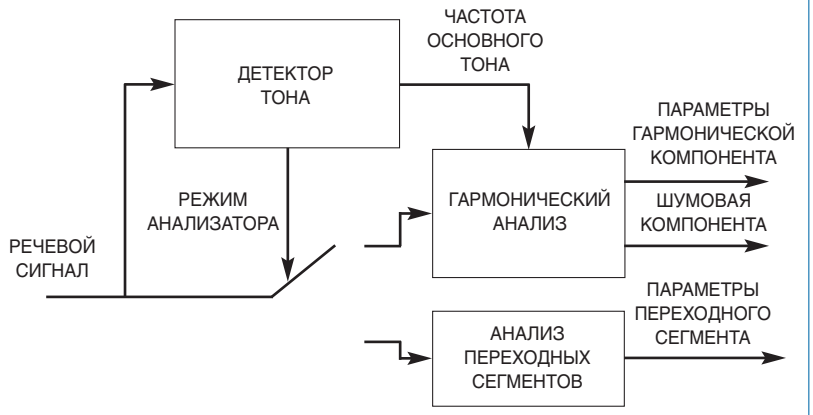

Рис. 2. Схема декомпозиции речевого сигнала

случаев стационарной вокализованной и стационарной невокализованной речи), с кодером формы сигнала, работающим во временной области (для переходных сегментов). Вторая работа использует сегментацию аудиосигнала на три различных сигнала: сигнал, моделирующий синусоидальную составляющую, сигнал, который моделирует все переходные сегменты, и шумовой сигнал.

Гармоническую составляющую речевого сигнала можно определить следующим образом:

$$h(i) = \sum_{k=1}^{K} A_k \cos\left(k \sum_{i=0}^{N-1} \frac{F_0(i)}{F_s} + \theta_k\right) \qquad (2)$$

где A_k — амплитуда k-ой гармоники, K — количество гармоник, $F_0(i)$ — мгновенная частота основного тона, θ_k — начальная фаза k-ой гармоники, F_s — частота дискретизации, N — длина сегмента.

Ядром гармонического анализа является процедура ДПФ, согласованного с изменением контура частоты основного тона (Pitch-Tracking Discrete Fourier Transform — PTDFT) [5]. Модифицированное ДПФ для анализа в области гармоник определяется так:

$$H_n(k) = \sum_{i=0}^{K-1} s_n(i) \exp\left(j \frac{2\pi k i}{F_s}\left(F_0 + \frac{\Delta F_0 i}{2N}\right)\right) w_n(i), \quad j = \sqrt{-1}, \qquad (3)$$

где $s_n(i)$ — i-й отсчёт n-го сегмента, F_0 — фундаментальная частота, ΔF_0 — приращение фундаментальной частоты, $w_n(i)$ — временное окно.

Таким образом, можно рассчитать амплитуды и фазы гармоник:

$$A_n(k) = \frac{\sqrt{\mathrm{Re}^2(H_n(k)) + \mathrm{Im}^2(H_n(k))}}{\sum_{i=0}^{L-1} w_n(i)},$$

$$\theta_n(k) = -arctg \frac{\mathrm{Im}(H_n(k))}{\mathrm{Re}(H_n(k))}.$$

Неортогональное ядро преобразования (3) может вызывать просачивание энергии в соседние спектральные отсчёты. Для устранения данного недостатка предлагается использовать времязависимое временное окно, форма которого пересчитывается синхронно с контуром изменения частоты основного

тона. Хорошие результаты получаются, если использовать в качестве прототипа окно Кайзера [13]:

$$w_n(i) = \frac{I_0\left(\beta\sqrt{1-\left[(2x-L_n+1)/(L_n-1)\right]^2}\right)}{I_0(\beta)},$$

где $i=0...N\text{-}1$, N — длина окна, β — параметр окна, $I_0(\ )$ — функция Бесселя нулевого порядка, x — функция, отражающая времязависимые характеристики:

$$x = \frac{a_{2,n}(N-1-i)^2 + a_{1,n}(N-1-i)}{a_{2,n}(N-1)+a_{1,n}},$$

где $a_{2,n}$ и $a_{1,n}$ — параметры, обеспечивающие линейное изменение фундаментальной частоты:

$$a_{2,n} = \frac{2\pi\Delta F_0}{F_s N},\ a_{1,n} = \frac{2\pi F_0}{F_s}.$$

2.2. Алгоритм определения параметров гармоник и частоты основного тона в цикле с обратной связью

Было показано [4, 5], что гармонический анализ с использованием PTDFT, а следовательно, и декомпозиция речевого сигнала являются достаточно корректными в случае точного определения значения частоты основного тона. Решение, которое предлагается здесь, подразумевает одновременное определение фундаментальной частоты и параметров (амплитуд и фаз) гармоник в цикле с обратной связью. Алгоритм грубой оценки частоты основного тона работает во временной области и основан на расчёте нормализованной автокорреляционной функции в комбинации с постобработкой на базе динамического программирования. В ходе предварительной оценки контура частоты основного тона производится низкочастотная фильтрация с частотой среза 1 кГц.

Информация о контуре фундаментальной частоты получается путём поиска максимумов нормализованной автокорреляционной функции (НАКФ):

$$\psi(k) = \frac{\sum_{j=1}^{N} s_j s_{j+k}}{\sqrt{\sum_{j=1}^{N} s_j^2 \sum_{j=1}^{N} s_{j+k}^2}},$$

где k — номер отсчёта НАКФ, соответствующий периоду основного тона. Максимумы, расположенные в пределах допустимых значений периода тона (в данном случае $16 \le k \le 160$), рассматриваются как кандидаты. Для того чтобы отбросить ложные пики, не рассматриваются кандидаты со значением автокорреляционной функции менее 30% от максимального на этом фрейме.

Следующий шаг алгоритма — отслеживание контура фундаментальной частоты, основанное на динамическом программировании (ДП) [14]. Так, для каждого кандидата рассчитывается

функция стоимости с учётом прошлой информации о контуре частоты основ ного тона: $D_{i,j} = d_{i,j} + \min_{k \in I_{i-1}} \{D_{i-1,k} + \delta_{i,j,k}\}$, где $d_{i,j}$ — локальная стоимость j-го кандидата в момент времени i, $\delta_{i,j,k}$ — стоимость перехода от k-го кандидата в момент времени i-1 к j-му кандидату в момент времени i ($1 \le j \le I_i$), I — количество кандидатов.

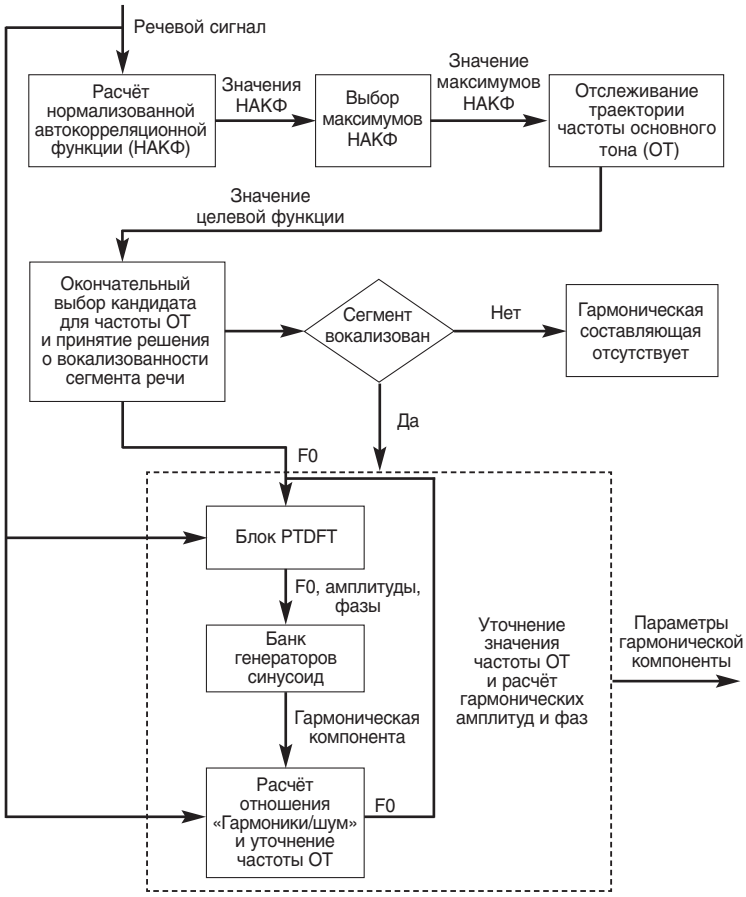

Рис. 3. Определение параметров гармоник и уточнение значения частоты основного тона в цикле с обратной связью

Целью данной процедуры является определение максимально гладкого контура частоты основного тона. После процедуры ДП в качестве предварительной оценки частоты основного тона на анализируемом сегменте выбирается кандидат j с минимальной стоимостью $D_{i,j}$. На ***рис. 3*** показана блок-схема алгоритма определения параметров гармоник и уточнения значения частоты основного тона в цикле с обратной связью.

Показателем точности определения параметров в данном случае может служить отношение «гармоники / шум»: $HNR = 10\lg(E_h/E^n)$, где E_H — энергия гармонической составляющей, синтезированной в соответствии с (2), E_H — энергия шумовой составляющей. Последняя определяется как разность между оригинальной речью и синтезированной гармонической компонентой: $r(i) = s(i) - h(i)$. Цикл с обратной связью для уточнения значения фундаментальной частоты выполняется после этапа грубой оценки. Целью данной процедуры является нахождение такого оптимального значения фундаментальной частоты, которое будет максимизировать значение HNR:

$$F_0^{opt} = arg\,max\,(\mathrm{HNR}(F_0)),\ F_{0min} \le F_0 \le F_{0max}.$$

Ядром данного процесса является PTDFT, а поиск оптимума осуществляется с помощью метода золотого сечения.

Для анализа переходных сегментов был использован подход ACELP в соответствии с рекомендациями ITU-T G.729 [15]. Пример декомпозиции речевого сигнала приведен на ***рис. 4***.

3. Конверсия голоса

3.1. Этап обучения

В работах [2, 16] была сделана попытка решения проблемы конверсии голоса с помощью таких методов преобразования спектра, как сопоставление кодовых книг амплитуд гармоник [16] и преобразование линейных спектральных частот (Line Spectral Frequencies — LSF) с помощью модели гауссовых смесей [2]. Принимая во внимание возможность использования модели (2), естественным было бы применить различные функции конверсии [17] для огибающих спектра каждой компоненты. При этом этап обучения осуществляется раздельно для составляющих модели (1).

На ***рис. 5*** показаны фазы обучения кодовых книг спектральных векторов.

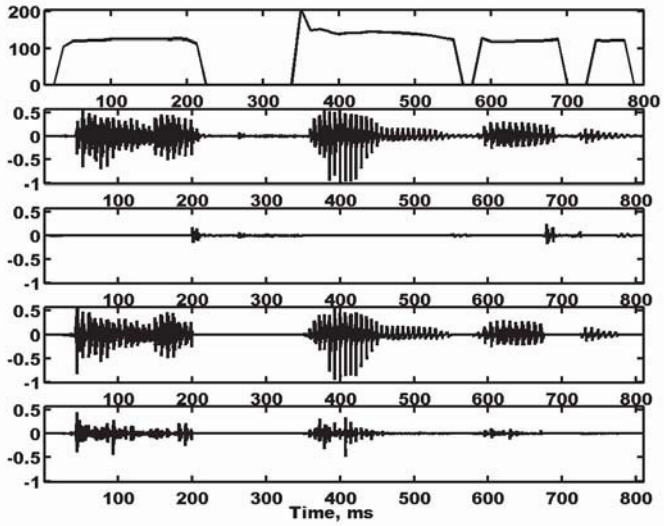

Рис. 4. Пример декомпозиции речевого сигнала. Сверху вниз: контур частоты основного тона, оригинальный речевой сигнал, переходные фреймы, гармоническая компонента, шумовая компонента

Из гармонического анализа для процесса конверсии берутся такие параметры, как спектральная огибающая, представленная LSF-коэффициентами, и фундаментальная частота F_0. Анализ переходных фреймов, осуществляемый с помощью метода ACELP [15], предоставляет для модификации LSF-коэффициенты фильтра, моделирующего вокальный тракт, период основного тона T_0, коэффициенты усиления последовательностей адаптивного и фиксированного возбуждения G_a и G_f соответственно.

Для осуществления преобразования таких параметров, как фундаментальная частота F_0, период основного тона T_0, коэффициенты усиления последовательностей адаптивного и фиксированного возбуждения G_a и G_f, использовался метод линейного преобразования математического ожидания и дисперсии. При этом предполагается, что математические ожидания этих параметров содержат существенную часть информации, специфической

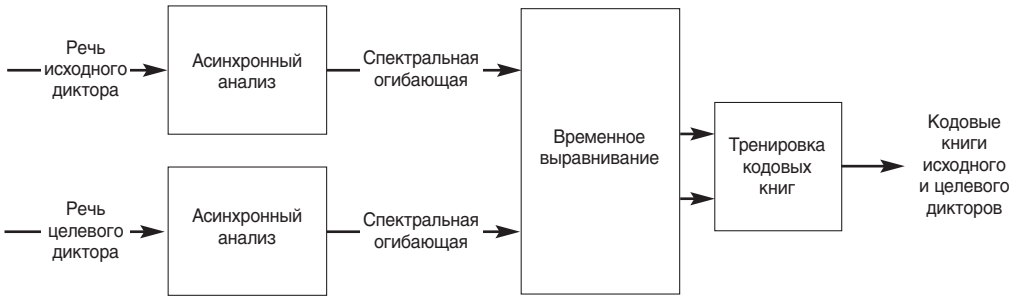

Рис. 5. Фазы этапа обучения кодовых книг

для каждого диктора. Предполагается также, что значения параметров каждого диктора подчиняются распределению Гаусса и имеют характерные средние значения и отклонения.

Обозначив модифицируемый параметр как $p_t^{S->T}$, можно определить линейное преобразование следующим образом [20]:

$$p_t^{S->T} = \frac{p_t^S - \mu^S}{\sigma^S}\sigma^T + \mu^T,$$

где p_t^T, p_t^S — один из параметров целевого и исходного дикторов соответственно, σ^S, μ^S, σ^T, μ^T — среднеквадратическое отклонение и математическое ожидание соответствующего параметра исходного и целевого дикторов соответственно.

В ходе обучения для установления более точного соответствия спектральных векторов исходного и целевого дикторов используется алгоритм динамической временной трансформации (DTW — Dynamic Time Warping) [20].

3.2. Процесс динамической временной трансформации

Предположим, имеется две последовательности наборов LSF-параметров: для целевого диктора — LSF^{tag} и для исходного — LSF^{source}. Необходимо выполнить выравнивание данных двух последовательностей по времени, т.е. таким образом сопоставить их по длине путём вставки или удаления элементов в LSF^{source}, чтобы общее среднеквадратическое отклонение между элементами этих последовательностей стремилось к минимальному значению.

Общее отклонение (расстояние) между последовательностями LSF^{tag} и LSF^{source} можно определить как $D\left(LSF^{tag}, LSF^{source}\right) = \sum_{s=1}^{Np} d\left(p_s\right)$, где $d(p_s)$ — расстояние между LSF_n^{source} и LSF_m^{tag}, N_p — количество элементов в пути. Тогда процесс нахождения оптимального пути сводится к минимизации общего отклонения: $P = arg\min_P \left(D\left(LSF^{tag}, LSF^{source}\right)\right)$.

Таким образом, задача сводится к нахождению функции выравнивания (пути):

$$P = p_1, \ldots, p_s, \ldots, p_{Np}, \quad p_s = (n_s, m_s),$$

каждое значение которой показывает, какой элемент последовательности LSF^{source} следует удалить, а какой вставить (***рис. 6***), чтобы достигнуть минимального значения общего отклонения.

Алгоритм для нахождения функции выравнивания (оптимального пути) может быть описан следующим образом:

Шаг 1. Вычислить матрицу локальных среднеквадратических отклонений d, каждый элемент которой является евклидовым расстоянием между двумя

соответствующими элементами последовательностей LSF^{source} и LSF^{tag} и определяется по следующему выражению:

$$d(n,m)=\sqrt{\sum_{k}^{10}\left(LSF_n^{source}(k)-LSF_m^{tag}(k)\right)^2},\ n=\overline{1,N},\ m=\overline{1,M}.$$

Шаг 2. Вычислить матрицу весов D, каждый элемент которой характеризует вклад соответствующего элемента матрицы d в общее среднеквадратическое отклонение.

Шаг 2.1. Положим начальное условие $D(1,1) = d(1,1)$. Вычислить первую строку матрицы D:

$$D(n,1)=D(n-1,1)+d(n,1),\ n=\overline{1,N}.$$

Шаг 2.2. Вычислить первый столбец матрицы D:

$$D(1,m)=D(1,m-1)+d(1,m),\ m=\overline{1,M}.$$

Шаг 2.3. Далее, двигаясь по матрице d слева направо снизу вверх, вычисляются следующие элементы матрицы D:

$$D(n,m)=\min\left[D(n,m-1),D(n-1,m-1),D(n-1,m)\right]+d(n,m),\ n=\overline{1,N},\ m=\overline{1,M}.$$

В процессе вычисления для каждой ячейки матрицы запоминается индекс соседней ячейки, которая вносит минимальный вклад в общую ошибку.

Шаг 3. Анализируя матрицу D в направлении от $D(N,M)$ до $D(1,1)$ и учитывая определённые на предыдущих этапах индексы ячеек, которые вносят меньший вклад в общее отклонение по сравнению с соседними, определяется наилучший путь $P = p_1, \ldots, p_k, \ldots, p_M$ с точки зрения минимизации величины общего отклонения.

Полученный в результате работы алгоритма путь $P = p_1, \ldots, p_k, \ldots, p_M$ является функцией сопоставления для обрабатываемых последовательностей, которая показывает, какие элементы необходимо удалить в исходной последовательности, а какие добавить.

Например, на ***рис. 7*** отображена функция сопоставления для двух наборов LSF-параметров исходного и целевого мужских дикторов с длиной $N=57$ и $M=68$ соответственно. Данным наборам соответствует фраза «Испорченный контакт».

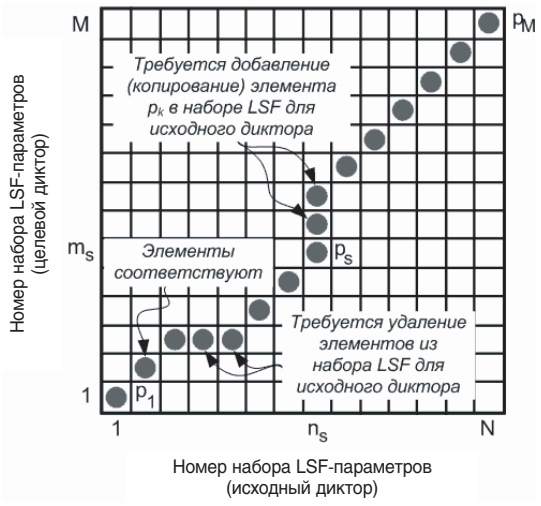

Рис. 6. Иллюстрация алгоритма динамической временной трансформации

3.3. Этап конверсии голоса

Функция конверсии векторов LSF для составляющих модели (1) имеет следующий вид:

$$F(x_t)=\sum_{i=1}^{N}p_i c_i,$$

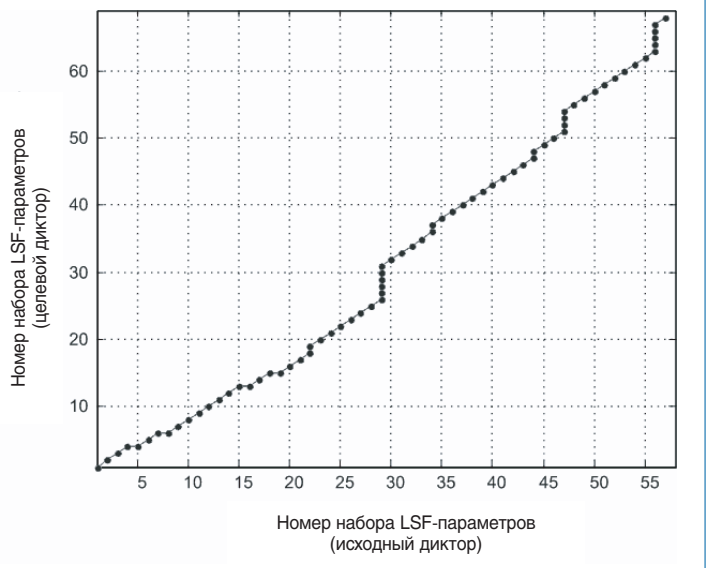

Рис. 7. Функция сопоставления для двух наборов LSF-параметров исходного и целевого мужских дикторов с длиной N=57 и M=68 соответственно

где p_i — вес, характеризующий вероятность принадлежности вектора x_t к i-му акустическому классу, представленному в кодовой книге размерности N центроидой c_i.

$$p_i = \frac{e^{-d_i}}{\sum_{j=1}^{N} e^{-d_j}},$$

где d_i — мера искажения: $d_i = \sum_{k=1}^{m} v_k |c_i - x_t|$.

Здесь величина m представляет собой размерность вектора, v_k — вес, рассчитанный по формуле обратного гармонического среднего, с помощью которого учитывается перцептуальный фактор близости смежных LSF:

$$v_k = \frac{1}{\omega_k - \omega_{k-1}} + \frac{1}{\omega_{k+1} - \omega_k},$$

где ω_k — k-й коэффициент LSF, ω_0=0, $\omega_{m+1} = \pi$.

На ***рис. 8*** показано, как выполняется процесс конверсии.

На вход системы поступает речевой сигнал, оцифрованный с частотой дискретизации 8 кГц. Детектор голосовой активности (VAD), реализованный в соответствии с [18], проверяет сегмент входного сигнала на наличие речи. Если

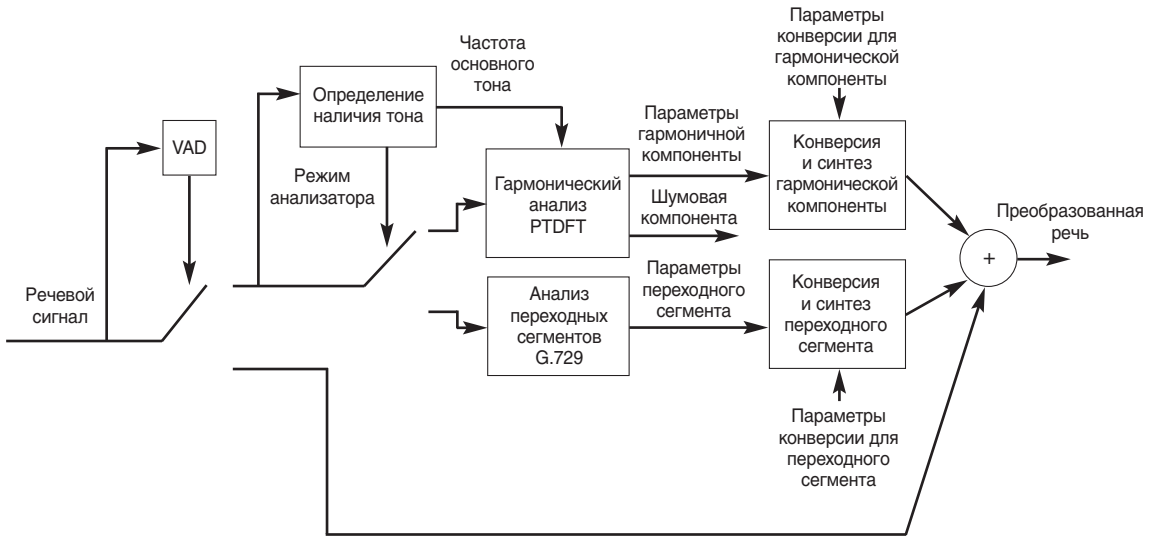

Рис. 8. Структурная схема процесса конверсии голоса

сегмент содержит тишину и/или фоновый шум, он передаётся напрямую на выход. Детектор тона определяет, будет ли речевой сегмент передан для обработки в модуль гармонического анализа либо в модуль анализа переходных сегментов. Параметры, выделенные с помощью одного из этих модулей анализа, подвергаются преобразованию, и далее осуществляется синтез фрагмента речи целевого диктора.

4. Экспериментальные результаты

Для сравнения использовалась система конверсии, полностью основанная на модели ACELP. Тесты показали, что голос, производимый предлагаемой системой, достаточно естественен и по разборчивости выше, чем в системе, основанной на подходе ACELP. Для примера была выбрана фраза на польском языке: «Lubić czardaszowy pląs» (***рис. 9***). 15 фраз из [19] были использованы для обучения.

Рис. 10 содержит спектрограммы примеров конверсии мужского голоса в женский. Очевидно, что результат конверсии предлагаемой системой лучше соответствует гармонической структуре речи целевого диктора и содержит меньше шума. Качество работы предлагаемой системы конверсии голоса оценивалось с помощью неформальных тестов прослушивания, которые показали, что узнаваемость диктора соответствует приблизительно 70%, реконструированная речь характеризуется достаточно высокой разборчивостью, хотя иногда характерны такие артефакты, как приглушённость и бормотание.

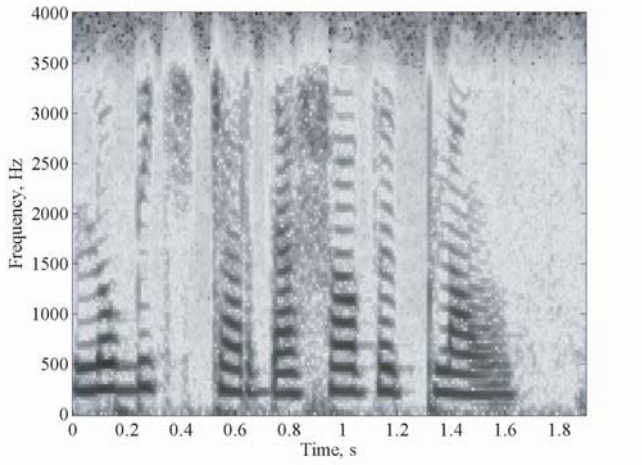

Рис. 9. Спектрограмма фразы целевого диктора

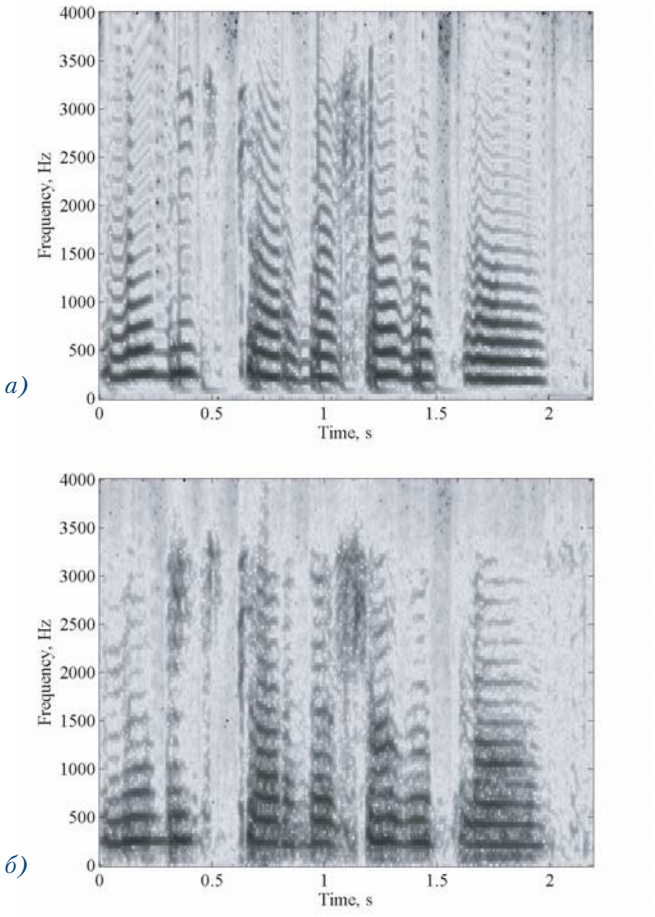

Рис. 10. Примеры конверсии голоса: а) системой, основанной на модели сепарации речевого сигнала; б) системой на базе ACELP-подхода

5. Заключение

В статье представлена система конверсии голоса, основанная на модели сепарации речевого сигнала на «гармоники+шум» и переходные фреймы с раздельной конверсией для каждой компоненты модели. Неформальные тесты прослушивания показали, что конвертированный речевой сигнал содержит небольшое количество артефактов, которые не мешают разборчивости и узнаваемости диктора.

Преимущество системы конверсии голоса в данном случае складывается из достоинств анализа-синтеза гармонической модели с достоинствами анализа и конверсии переходных сегментов во временной области. Благодаря данному подходу, значительно снижается доля исходного диктора в конвертированной речи, по сравнению, например, с системой конверсии на базе ACELP подхода.

Литература

1. *Moulines E. and Sagisaka Y., Eds.* «Voice conversion: state of the art and perspectives». Speech Communication, vol. 16, Feb. 1995.

2. *Pavlovets A., Kien T., Zubricki P. and Petrovsky A.* «Speech analysis — synthesis based on the PTDFT for voice conversion», in Proc. of the 2007 Int. Workshop on Spectral Methods and Multirate Sig. Proc., SMMSP, Moscow, Russia, Sep. 2007, pp. 203–210.

3. *Stylianou Y., Laroche J. and Moulines E.* «High-quality speech modification based on a harmonic + noise model», in Proc. of the European Conf. on Speech Communication and Technology EUROSPEECH, Madrid, Spain, Sep. 1995, pp. 451–454.

4. *Petrowsky A., Zubricki P. and Sawicki A.* «Tonal and noise components separation based on a pitch synchronous DFT analyzer as a speech coding method,» in Proc. European Conf. Circuit Theory and Design, Cracow, Poland, Sep. 2003, vol. 3, pp.169–172.

5. *Zubricki P., Pavlovec A. and Petrovsky A.* «Analysis-by-synthesis parameters estimation in the harmonic coding framework by pitch tracking modified DFT» in «New trends in audio and video», Dobrucki A., Petrovsky A. and Skarbek W. Eds. Byalystok 2006, pp. 233–246.

6. *Tremain T.* «The government standard linear predictive coding algorithm: LPC-10», Speech Technology Magazine, vol. 1, №. 2, 1982, pp. 40–49.

7. *Griffin D. and Lim J.* «Multiband excitation vocoder», IEEE Trans. Acoust., Speech and Sig. Proc., vol. 36, №8, pp. 1223 — 1235, Aug. 1988.

8. *Yegnanarayana B., d'Alessandro C. and Darsinos V.* «An iterative algorithm for decomposition of speech signals into periodic and aperiodic components», IEEE Trans. on Speech and Audio Proc., vol.6, № 1, pp. 1–11, Jan. 1998.

9. *Jackson P.J.B. and Shadle C.H.* «Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech», IEEE Trans. on Speech and Audio Proc., vol.9, №7, pp.713–726, Oct.2001.

10. *Abe M., Nakamura S., Shikano K. and Kuwabara H.* «Voice conversion through vector quantization», in Proc. of the Int. Conf. on Acoust., Speech and Sig. Proc. ICASSP, New York, USA, Apr.1988, vol.1, pp.655–658.

11. *Shlomot E., Cuperman V. and Gersho A.* «Hybrid coding: combined harmonic and waveform coding of speech at 4 kb/s», IEEE Trans. on Speech and Audio Proc., vol.9, № 6, pp. 632–646, Sep. 2001.

12. *Levine S. and Smith J.O.* «A sines+transients+noise audio representation for data compression and time/pitch scale modifications» in Proc. 105th Conv. Audio Eng. Soc., preprint #4781, Sep.1998.

13. *Sercov V. and Petrovsky A.* «An improved speech model with allowance for time-varying pitch harmonic amplitudes and frequencies in low bit-rate MBE coders», in Proc. of the European Conf. on Speech Communication and Technology EUROSPEECH, Budapest, Hungary, Sep.1999, pp.1479 — 1482.
14. *Talkin D.* «Robust algorithm for pitch tracking» in «Speech Coding and Synthesis», Kleijn W.B. and Palival K.K. Eds. Elsevier, Amsterdam, Netherlands, 1995.
15. ITU-T Rec. G.729, «Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear — prediction (CS-ACELP)», Mar.1996.
16. *Pavlovets A. and Petrovsky A.* «Voice conversion as a part of the voice analysis/synthesis system based on the periodic-aperiodic decomposition of speech», in Proc. of the 9th Int. Conf. on Pattern Recognition and Information Proc., PRIP, Minsk, Belarus, May2007, vol.2, pp. 71–76.
17. *Stylianou Y., Cappe O., Moulines E.* «Continuous probabilistic transform for voice conversion», IEEE Trans. on Speech and Audio Processing, vol. 6, № 2, pp. 131–142, March 1998.
18. ITU-T Rec. G.729, annex B, «A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70», Nov.1996.
19. *Grocholevski S.* «First database for spoken Polish», in Proc. Int. Conf. On Language Resources and Evaluation, Grenada, 1998, pp. 1059–1062.
20. *Huang X., Acero A., Hon H-W.* «Spoken Language Processing: a guide to theory, algorithms, and system development», Prentice Hall, NJ, 2001. — 980 p.

Павловец Александр Николаевич —

аспирант-заочник в Учреждении образования «Белорусский государственный университет информатики и радиоэлектроники». Закончил Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» по специальности «Проектирование и технология электронных вычислительных средств». Работает на Заводе вычислительной техники им. С. Орджоникидзе. Область интересов — цифровая обработка речевых сигналов, кодирование речевых сигналов, проектирование проблемно-ориентированных средств вычислительной техники реального времени для мультимедиа-систем.

Лившиц Михаил Зеннадьевич —

кандидат технических наук, доцент кафедры Электронных вычислительных средств Учреждения образования «Белорусский государственный университет информатики и радиоэлектроники». Закончил Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» по специальности «Электронные вычислительные средства». Область интересов — разработка реконфигурируемых аппаратных платформ для мультимедиа-систем, цифровая обработка сигналов, кодирование широкополосных речевых и аудиосигналов, повышение качества речевых сигналов, конверсия голоса.

Лихачёв Денис Сергеевич —

кандидат технических наук, доцент кафедры Электронных вычислительных средств Учреждения образования «Белорусский государственный университет информатики и радиоэлектроники». Закончил Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» по специальности «Проектирование и технология электронных вычислительных средств». Область интересов — цифровая обработка речевых сигналов, системы компрессии речи, антропоморфическая обработка речи, конверсия голоса.