

ИЗМЕНЕНИЕ ОСНОВНОГО ТОНА РЕЧЕВОГО СИГНАЛА В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ

И.С. АЗАРОВ¹, М.И. ВАШКЕВИЧ², Д.С. ЛИХАЧЕВ³, А.А. ПЕТРОВСКИЙ⁴

*Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, г. Минск, 220013, Республика Беларусь
¹azarov@bsuir.by, ²vashkevich@bsuir.by, ³likhachov@bsuir.by, ⁴palex@bsuir.by*

Рассматривается задача изменения частоты основного тона речевого сигнала. Предложенное решение позволяет выполнять обработку широкополосного речевого сигнала в реальном масштабе времени. Допускается изменение исходного контура частоты основного тона на произвольный целевой. Выполняется оценка качества обработки сигнала.

Ключевые слова: оценка частоты основного тона, моделирование речевого сигнала.

В разработанной системе изменения основного тона используется синусоидальная модель речевого сигнала: $s(n) = \sum_{k=1}^K A_k(n) \cos \varphi_k(n) + r(n)$, где $A_k(n)$ – мгновенная амплитуда k -ой гармоники, K – число гармоник, $\varphi_k(n)$ – мгновенное значение фазы k -ой гармоники, $r(n)$ – шумовая составляющая сигнала [1]. Мгновенная частота $f_k(n)$ связана с мгновенной фазой следующим соотношением: $\varphi_k(n) = \sum_{l=0}^n \frac{2\pi f_k(l)}{F_s} + \varphi_k(0)$, где F_s – частота дискретизации и $\varphi_k(0)$ – начальная фаза k -ой гармоники. Приближенно можно считать, что частота каждой гармоники является кратной частоте основного тона т.е. $f_k(n) \approx F_0(n)k$, где $F_0(n)$ – основной тон.

Частота основного тона определяет высоту звучания голоса. Контур частоты основного тона (его изменение в зависимости от времени) определяет интонацию речи. Для того чтобы сохранить исходный тембр диктора должна сохраняться спектральная огибающая, которую можно определить из мгновенных амплитуд гармоник основного тона. Огибающую можно рассматривать как функцию от номера отсчета и частоты $E(n, f)$, которая принимает значения мгновенных амплитуд гармоник основного тона в соответствующих точках $E(n, f_k(n)) = A_k(n)$. Для произвольных n и f функция вычисляется путем линейной интерполяции ближайших к ним амплитудных значений. Синтез голоса с модифицированным контуром частоты основного тона может быть выполнен по следующей формуле: $s(n) = \sum_{k=1}^K E(n, F_0(n)k) \cos \bar{\varphi}_k(n) + r(n)$, где фазы гармонических компонент $\bar{\varphi}_k(n)$ рассчитываются в соответствии с новым контуром частоты основного тона $F_0(n)$ как $\bar{\varphi}_k(n) = \sum_{l=0}^n \frac{2\pi F_0(l)}{F_s} + \bar{\varphi}_k^{\Delta}(n)$. Дополнительный фазовый параметр $\bar{\varphi}_k^{\Delta}(n)$ используется для сохранения относительных фаз гармоник по отношению к фазе частоты основного тона. Данный параметр вычисляется как $\bar{\varphi}_k^{\Delta}(n) = \varphi_k(n) - k\varphi_0(n)$.

Основной тон присутствует только в вокализованных сегментах речи (там где задействованы голосовые связки диктора). Такие звуки как 'а', 'о', 'э' являются вокализованными, в то время как звуки 'с', 'ш', 'щ' являются невокализованными. Наличие вокализованности проявляется в виде спектральных компонент кратной частоты. Для того, чтобы сохранить исходное качество звучания невокализованных звуков алгоритм изменения тона автоматически выделяет области вокализованности и выполняет обработку только в этих областях. Для этого шумовая часть сигнала $r(n)$ выделяется из исходного сигнала вычитанием выделенных вокализованных звуков.

Алгоритм обработки сигнала состоит из следующих основных шагов (рисунок 1): 1) Определение частоты основного тона $F_0(n)$ [2]; 2) Преобразование речевого сигнала в параметрический вид т.е. оценка мгновенных гармонических параметров $A_k(n)$, $f_k(n)$ и $\varphi_k(n)$, $k = 1, 2, \dots, K$; 3) Оценка вокализации каждой тройки гармонических параметров и отбор только тех, которые относятся к вокализованным областям спектра; 4) Синтез исходной вокализованной компоненты сигнала и ее вычитание из исходного речевого сигнала для получения шумовой составляющей $r(n)$; 5) Синтез вокализованной компоненты с измененным основным тоном и ее сложение с шумовой составляющей $r(n)$.

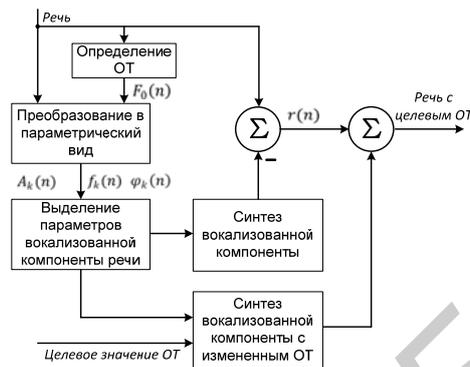


Рис. 1. Алгоритм изменения основного тона

Алгоритм реализован в виде независимых функциональных блоков, которые принимают входные сигналы и формируют выходные в соответствии с заданной функцией преобразования. Передача сигнала между блоками выполняется при помощи линий сигнала, которые представляют собой массивы данных с изменяющимися размерами. После выполнения требуемого преобразования входных сигналов каждый блок обновляет входные и выходные линии, удаляя данные из входных линий и добавляя данные к выходным. В случае, если на входе блока недостаточно данных для выполнения преобразования, то он не выполняет никаких действий. Такая схема реализации обусловлена необходимостью реализации в системе многоскоростной обработки, не позволяющей заранее организовать жесткую синхронизацию внутренних сигналов.

Тестирование реализованной системы выполнялось на речевой базе, включающей 9 различных голосов (4 мужских голоса и 5 женских). Длительность записи каждого диктора приблизительно составляла 1 мин 10 секунд. Чтобы оценить качество работы в телекоммуникационной системе наряду с исходными сигналами обрабатывались сигналы закодированные кодеками GSM и G.711.

Список литературы

1. Abe, T. and Honda, M. "Sinusoidal model based on instantaneous frequency attractors", IEEE Transactions on Audio, Speech, and Language Processing, Volume: 14, Issue 4, pp. 1292 – 1300, July 2006.
2. Azarov, E. Instantaneous Pitch Estimation Based on RAPT Framework / E. Azarov, M. Vashkevich and A. Petrovsky // Proc. of the 20th European Signal Process. Conf. (EUSIPCO-2012). – 2012. – P. 2787–2791.