

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ИЗМЕНЕНИЯ ВЫСОТЫ ЗВУЧАНИЯ РЕЧИ С СОХРАНЕНИЕМ ТЕМПА ВОСПРОИЗВЕДЕНИЯ

Жиляков Е.Г., Фирсова А.А., Курлов А.В.

Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

С развитием компьютерных технологий для большинства пользователей стало возможно записывать звуковой сигнал и осуществлять его редактирование и обработку. В ряде случаев, возникает необходимость изменение высоты голоса (звуковысотности) с сохранением темпа воспроизведения [1].

В основе предлагаемого метода используется вычисление субполосных огибающих в заданных частотных интервалах [2]. При этом ось нормированных частот $[0, \pi]$ разбивается на ряд неперекрывающихся частотных интервалов V_r , $r = 0, \dots, R$, границы которых определяются соотношениями:

$$V_{10} = 0; V_{20} = \pi/(N-1) = \pi/(2R+1); V_{1r} = V_{2,r-1}; V_{2r} = V_{1r} + 2\pi/(N-1);$$

$$\omega_0 = \pi/(N-1), \omega_r = r \cdot 2\omega_0, r = 1, \dots, R; R = (N-2)/2,$$

где N – длительность анализируемого отрезка, V_{10} и V_{20} – границы нулевого частотного интервала. Пусть \vec{x}_N – вектор значений исходного отрезка речевого сигнала. Тогда результат преобразования можно записать как:

$$\widehat{\vec{x}}_N = \sum_{r=0}^R (C_r^+ \vec{v}_c^r + S_r^+ \vec{v}_s^r), \quad \text{где} \quad \vec{v}_c^r = A_0 C_r \vec{x}_N, \quad \vec{v}_s^r = A_0 S_r \vec{x}_N \quad - \quad \text{огibaющие}$$

соответствующих косинусоид и синусоид (диагоналей матриц $C_r = \text{diag}(1, \cos(\omega_r), \dots, \cos(\omega_r(N-1)))$ и $S_r = \text{diag}(0, \sin(\omega_r), \dots, \sin(\omega_r(N-1)))$);

$A_0 = \{a_{ik}^0\}$ – субполосная матрица с элементами $a_{ik}^0 = \sin(\omega_0(i-k)) / \pi(i-k)$, $a_{ii}^0 = \omega_0 / \pi$, $i, k = 1, \dots, N$; $C_r^+ = \text{diag}(1, \cos(\omega_r + \Delta_r), \dots, \cos((\omega_r + \Delta_r) \cdot (N-1)))$, $S_r^+ = \text{diag}(0, \sin(\omega_r + \Delta_r), \dots, \sin((\omega_r + \Delta_r) \cdot (N-1)))$, Δ_r – некоторое желаемое смещение (с учетом знака) центра r -го частотного интервала.

Необходимо отметить, что при таком преобразовании не изменяется темп (скорость воспроизведения), а изменяется только высота звучания. Так, выбор смещения Δ_r со знаком плюс, приводит к смещению спектра в область высоких частот. Выбор же смещения Δ_r со знаком минус приводит к смещению спектра в низкочастотную область. Результаты экспериментов с реальными речевыми сигналами, записанными в режиме моно с частотой дискретизации 11025 Гц, разрядностью кода 16 бит (при $N = 258$), подтвердили правильность полученных преобразований.

Литература:

1. Ширяев А. Методы транспозиции музыкального материала в звукорежиссуре / А. Ширяев // Звукорежиссер. – 2011. – №6. – С. 57-60.
2. Жиляков Е.Г. Компьютерная транспозиция спектра речевых сигналов / Е.Г. Жиляков, А.В. Курлов, А.А. Фирсова // Вопросы радиоэлектроники, сер. ЭВТ. – Выпуск 2. – 2014. – С. 160-172.

Работа выполнена при поддержке гранта РФФИ 12-07-00514А.