

ПЕРЕДИСКРЕТИЗАЦИЯ РЕЧЕВОГО СИГНАЛА СОГЛАСОВАННАЯ С ЧАСТОТОЙ ОСНОВНОГО ТОНА

М.И. ВАШКЕВИЧ¹, И.С. АЗАРОВ², Д.С. ЛИХАЧЕВ³, А.А. ПЕТРОВСКИЙ⁴

*Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, г. Минск, 220013, Республика Беларусь*

vashkevich@bsuir.by, azarov@bsuir.by, likhachov@bsuir.by, palex@bsuir.by

Рассматривается задача согласованной с основным тоном передискретизации речевого сигнала. Предложенное решение позволяет выполнять динамическое масштабирование временной оси таким образом, чтобы на каждый период частоты основного тона речевого сигнала приходилось равное число временных отсчетов. Приводятся результаты экспериментов.

Ключевые слова: частота основного тона, обработка речевых сигналов.

Речевой сигнал представляет собой сложный нестационарный процесс. Важной его отличительной чертой является наличие в нем фрагментов, имеющих гармоническую структуру. Они отвечают за вокализованные звуки (например "о", "а", "э" и проч.). Для многих вокализованных звуков характерна нестабильность частоты основного тона. Вследствие чего спектральный анализ таких звуков делается затруднительным. Особенно остро эта проблема встает в тех случаях, когда необходимо выполнить оценку параметров отдельных гармоник вокализованного сигнала. В частности такая задача возникает при использовании синусоидальной модели 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: $\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0)$, где F_s – частота дискретизации и $\varphi_k(0)$ – начальная фаза k -ой гармоники. Приблизительно 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.

В данной работе для упрощения задачи оценки параметров отдельных гармоник вокализованного сигнала предлагается способ передискретизации речевого сигнала согласованно с частотой основного тона. Делается допущения, что для вокализованного речевого сигнала $s(n)$ известен контур частоты основного тона $f_0(n)$. Целью передискретизации сигнала является перерасчет его в новые моменты времени m таким образом, чтобы на каждый период основного тона приходилось равное количество отсчетов N_{f_0} . Временные метки m отстоят друг от друга на неравные интервалы, которые зависят от частоты основного тона.

Каждому входному отсчету речевого сигнала $s(n)$ ставится в соответствие фазовую отметку $p = \phi(n) = \sum_{i=0}^n f_0(i)$, где $f_0(i)$ – нормализованная круговая частота основного тона в момент времени i . Новые временные моменты m , в которые необходимо перерасчитать входной сигнал, определяются как

$$m = \phi^{-1}(q/N_{f_0}),$$

где q – это индекс отсчета в деформированной временной области. Рис. 1,а поясняет приведенную идею вычисления $\phi^{-1}(\cdot)$.

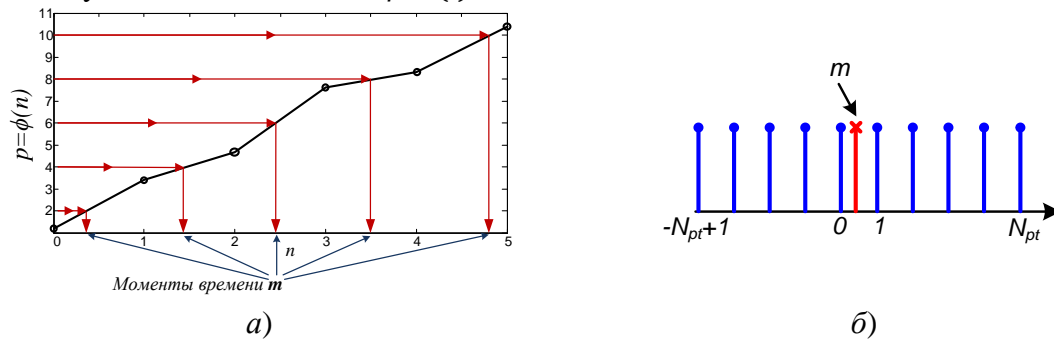


Рисунок 1 – Расчет моментов времени m : а) вычисление моментов m ; б) нормировка

Восстановление значения сигнала в заданный момент времени выполняется с использованием теоремы Котельникова, согласно которой: $s(m) = \sum_{n=-\infty}^{\infty} s(nT) \text{sinc}\left(\frac{m-nT}{T}\right)$, где $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$, T – интервал дискретизации, а $s(nT)$ – значения сигнала в дискретные моменты времени nT . Для простоты примем $T = 1$, тогда $s(nT) = s(n)$. Для практического использования в выражении для $s(m)$ ограничивают пределы суммирования, выбирая N_{pt} точек сигнала предшествующих моменту m и N_{pt} последующих: $s(m) = \sum_{n=-N_{pt}}^{N_{pt}} s(n) \text{sinc}(m - n)$. Для каждого вычисления $s(m)$ временные отметки пересчитываются таким образом, чтобы момент m попадал в диапазон от 0 до 1. Это позволяет подготовить таблицу с заранее рассчитанными значениями функции sinc для разных смещений в диапазоне от 0 до 1. Использование таблицы дает возможность избежать многократного вычисления функции синус в процессе вычисления. Результат работы предложенного способа показан на рис. 2.

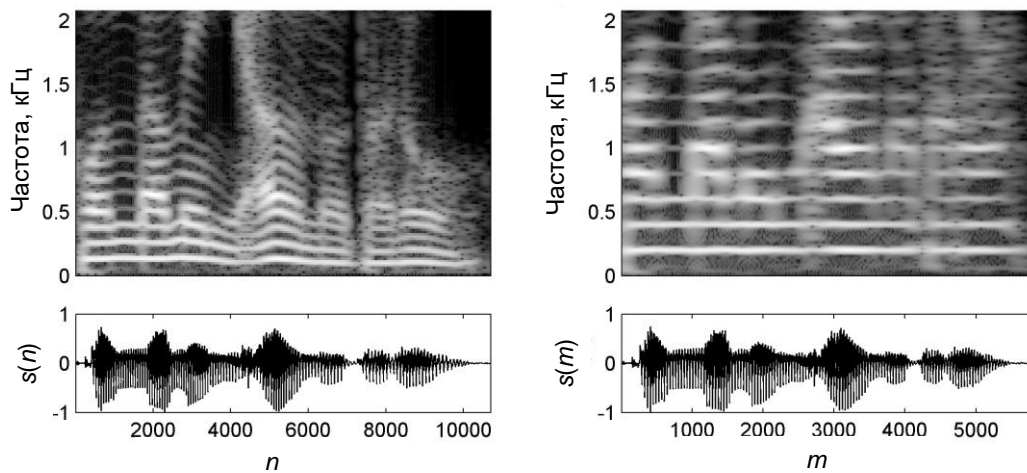


Рисунок 2 – Результат передискретизации согласованной с частотой основного тона

Сигнал, получаемый после передискретизации имеет постоянную частоту основного тона. Вследствие этого процедура оценка параметров отдельных гармоник сигнала может быть значительно упрощена.

Список источников

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.