

ПЕРЕДИСКРЕТИЗАЦИЯ РЕЧЕВОГО СИГНАЛА СОГЛАСОВАННАЯ С ЧАСТОТОЙ ОСНОВНОГО ТОНА

М.И. Вашкевич, И.С. Азаров, Д.С. Лихачев
А.А. Петровский

Кафедра электронных вычислительных средств,
Белорусский государственный университет информатики и
радиоэлектроники,
Минск, Беларусь

1. Постановка задачи

Рассматривается задача согласованной с основным тоном передискретизации речевого сигнала, которая заключается в динамическом масштабировании временной оси таким образом, чтобы на каждый период частоты основного тона речевого сигнала приходилось равное число временных отсчетов.

Предпосылки:

- Речь представляет собой нестационарный сигнал, имеющий гармоническую структуру;
- Большинство известных методов анализа предназначены для стационарных сигналов;

2. Модель речевого сигнала

Синусоидальная модель речевого сигнала

$$s(n) = \sum_{k=1}^K A_k(n) \cos \varphi_k(n) + r(n),$$

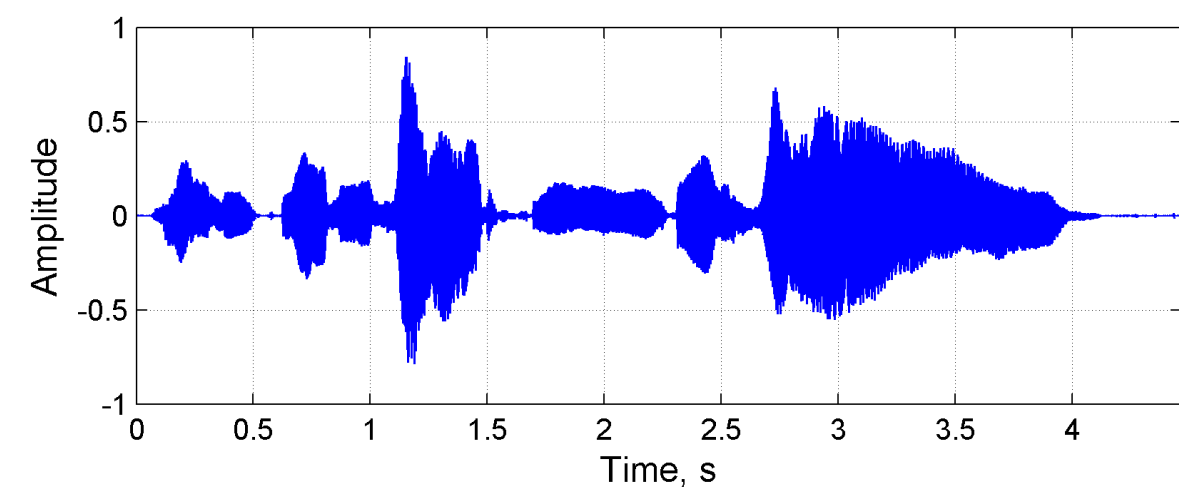
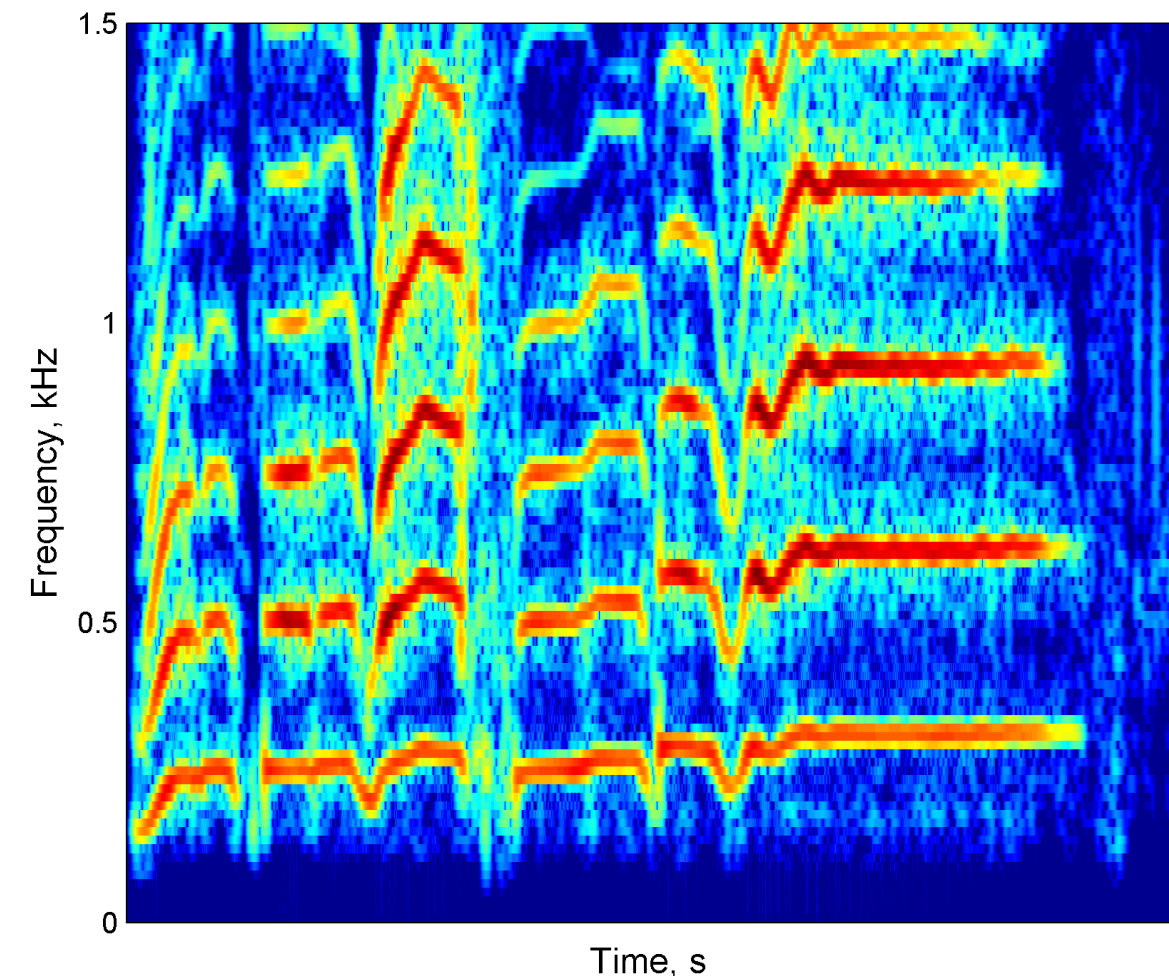
где $A_k(n)$ и $\varphi_k(n)$ – мгновенная амплитуда и фаза k -ой гармоники, K – число гармоник, $r(n)$ – шумовая составляющая сигнала.

Мгновенная частота $f_k(n)$ связана с мгновенной фазой:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0),$$

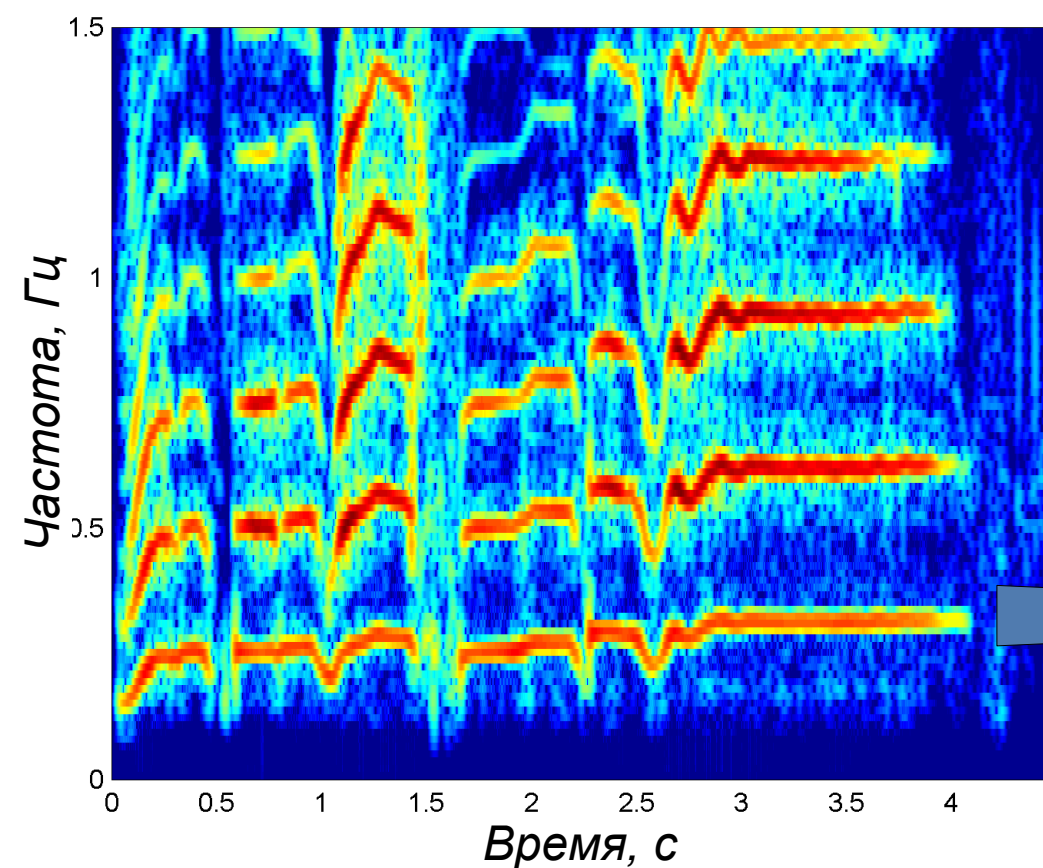
где F_s – частота дискретизации и $\varphi_k(0)$ – начальная фаза k -ой гармоники. Приблизительно можно считать, что частота гармоники является кратной частоте основного тона:

$f_k(n) \approx F_0(n)k$, где $F_0(n)$ - **основной тон**.

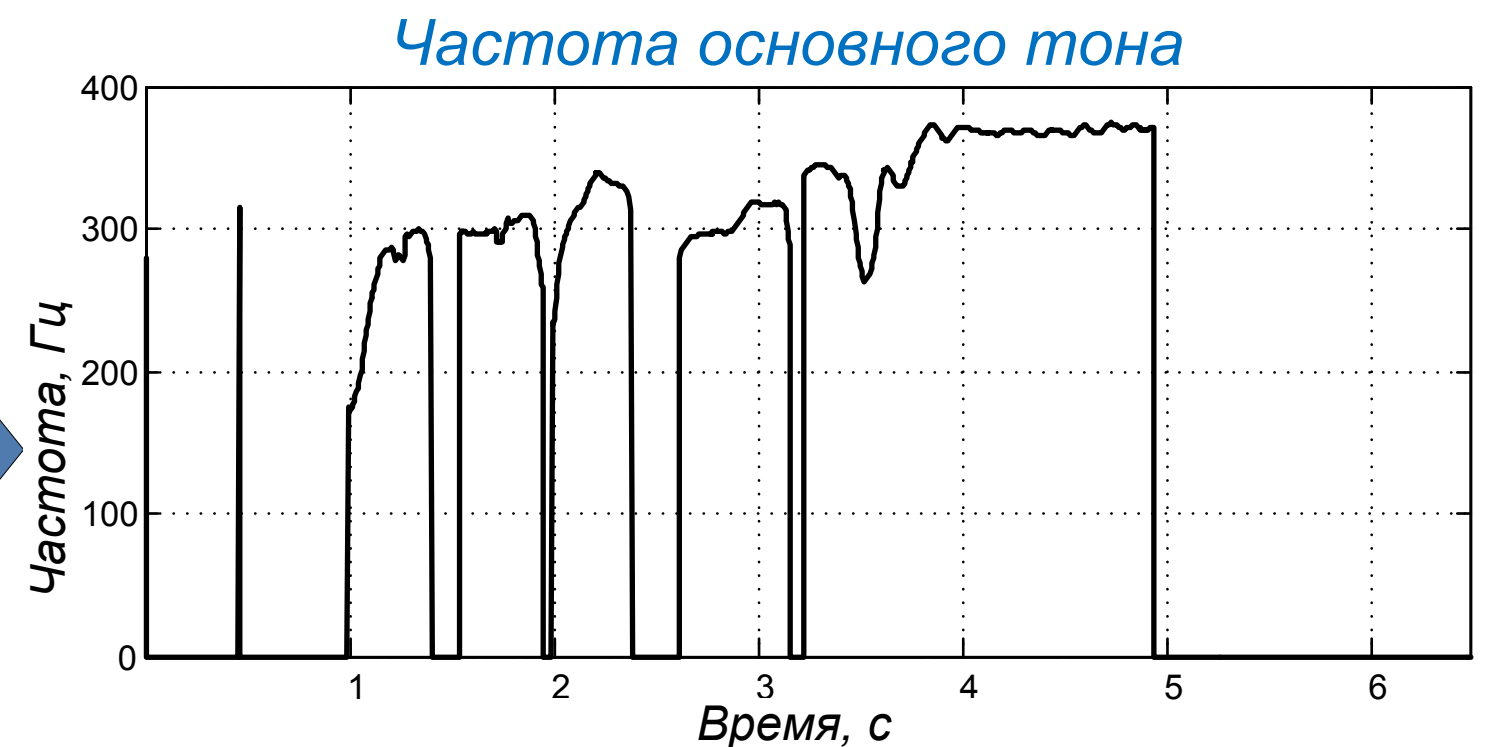


3. Выделение частоты основного тона

Выделение частоты основного тона (ЧОТ) выполняется на основе модификации алгоритма **RAPT** (robust algorithm for pitch tracking) – **IRAPT**, в котором автокорреляционная функция рассчитывается из параметрической формы представления речевого сигнала.



IRAPT

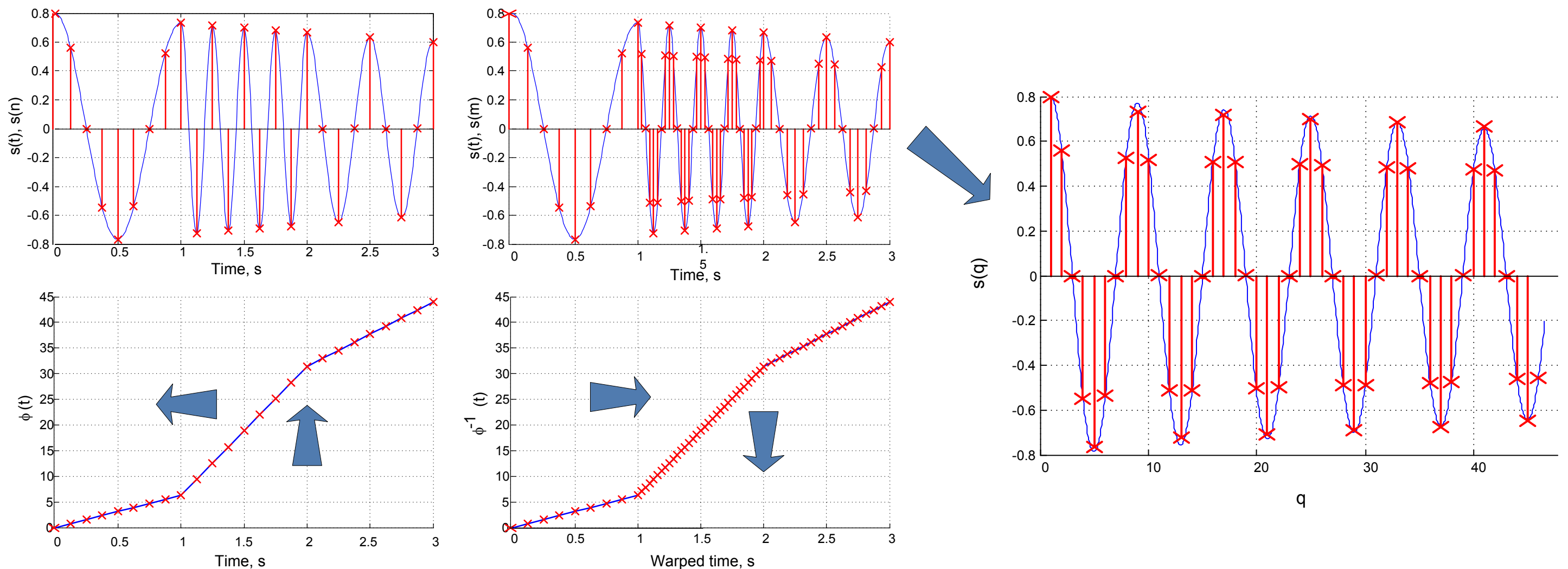


4. Передискретизация речевого сигнала

Целью передискретизации сигнала является перерасчет его в новые моменты времени t таким образом, чтобы **на каждый период основного тона приходилось равное количество отсчетов**.

Каждому отсчету сигнала $s(n)$ ставится в соответствие фазовая метка $\phi(n) = \sum_{i=0}^n f_0(i)$, где $f_0(i)$ – ЧОТ в момент времени i . Временные моменты t , определяются как

$t = \phi^{-1}(q/N_{f_0})$, где q – это индекс отсчета в деформированной временной области.



5. Перерасчет сигнала

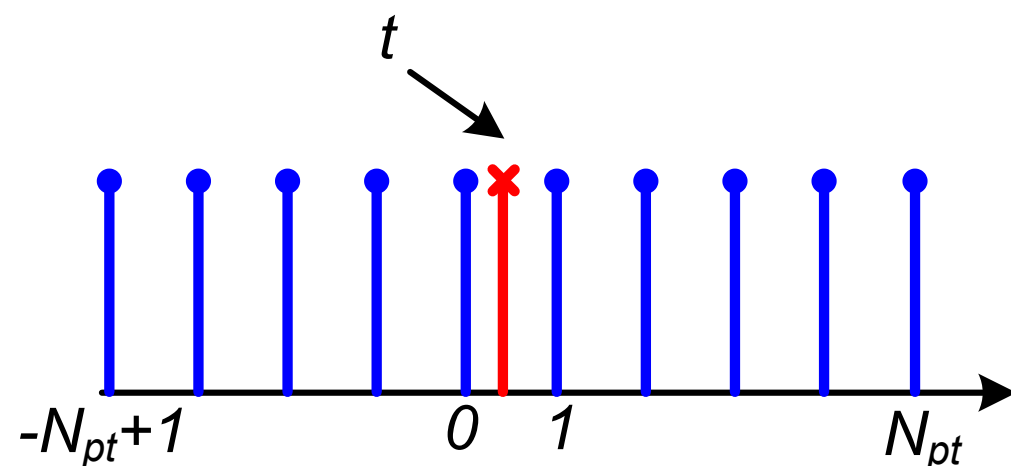
Согласно теореме Котельникова значения сигнала в момент времени t рассчитывается как:

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \operatorname{sinc}\left(\frac{t - nT}{T}\right),$$

где $\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x}$, T – интервал дискретизации, а $s(nT)$ – значения сигнала в дискретные моменты времени nT . Для простоты примем $T = 1$.

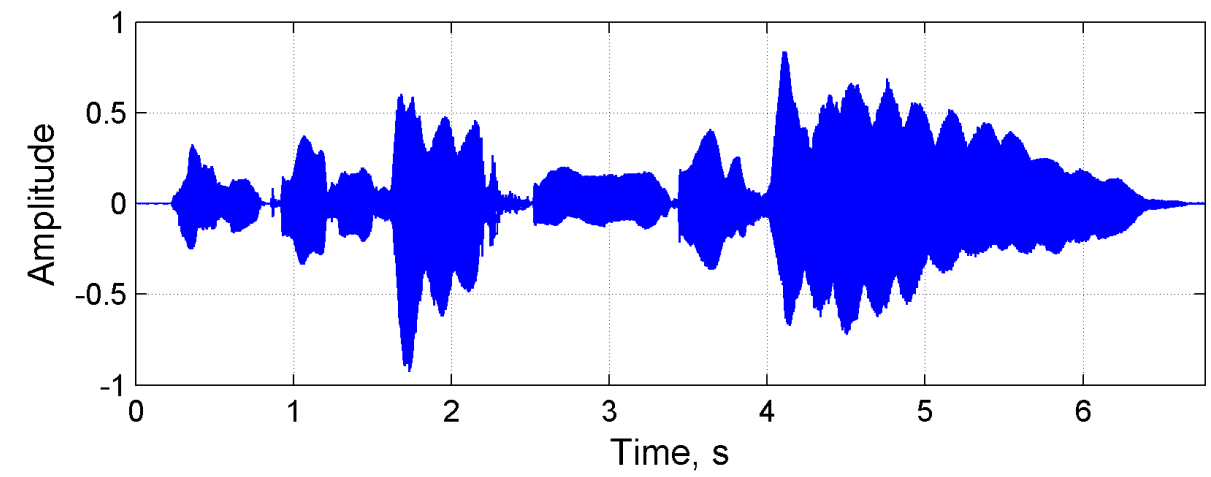
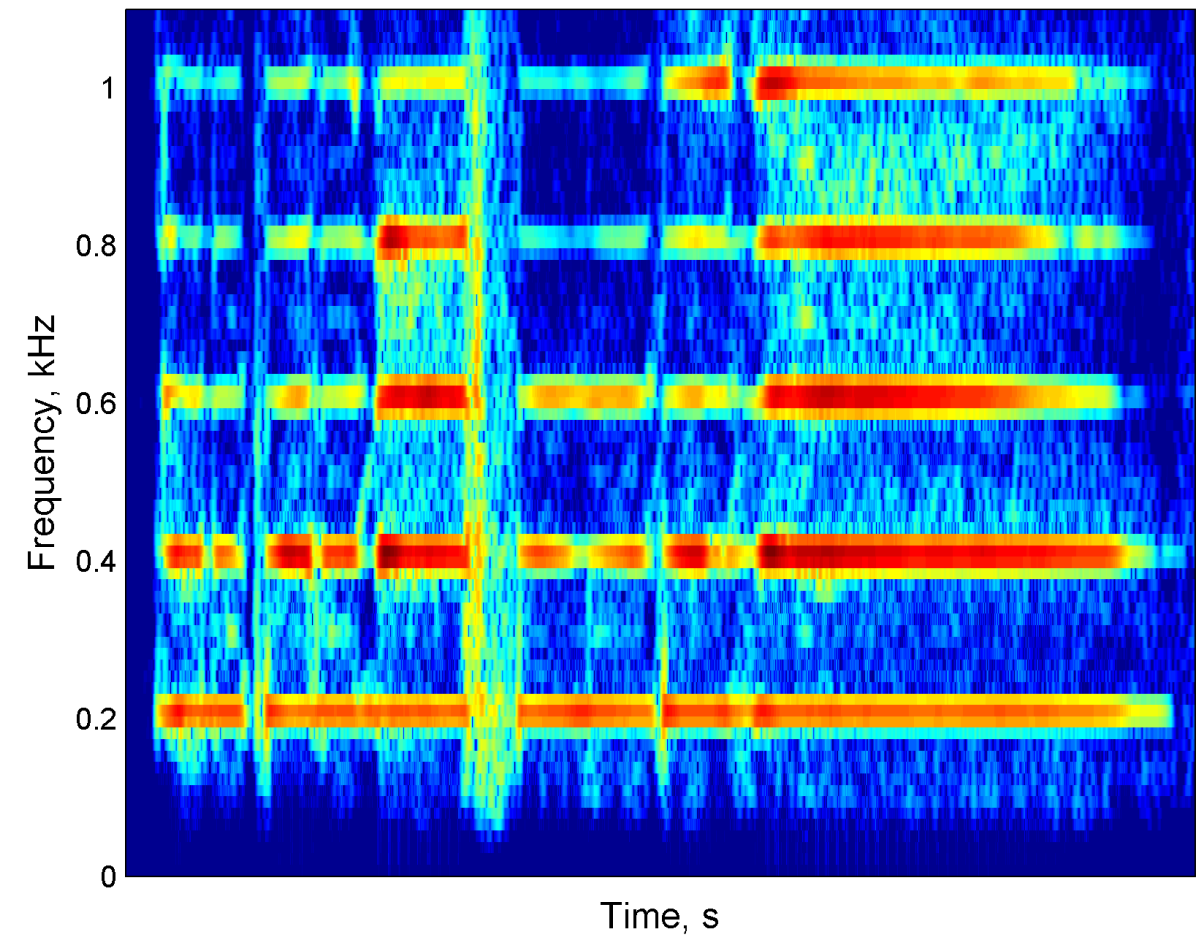
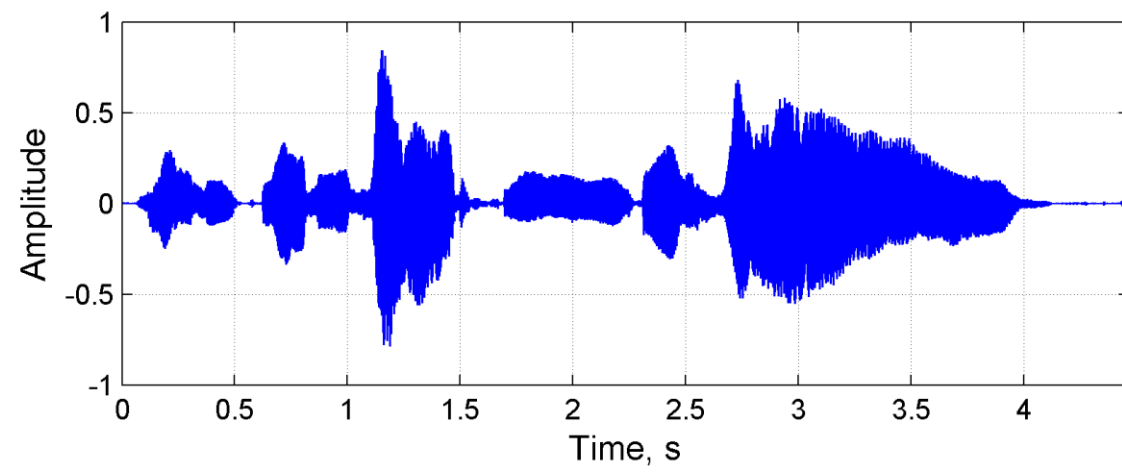
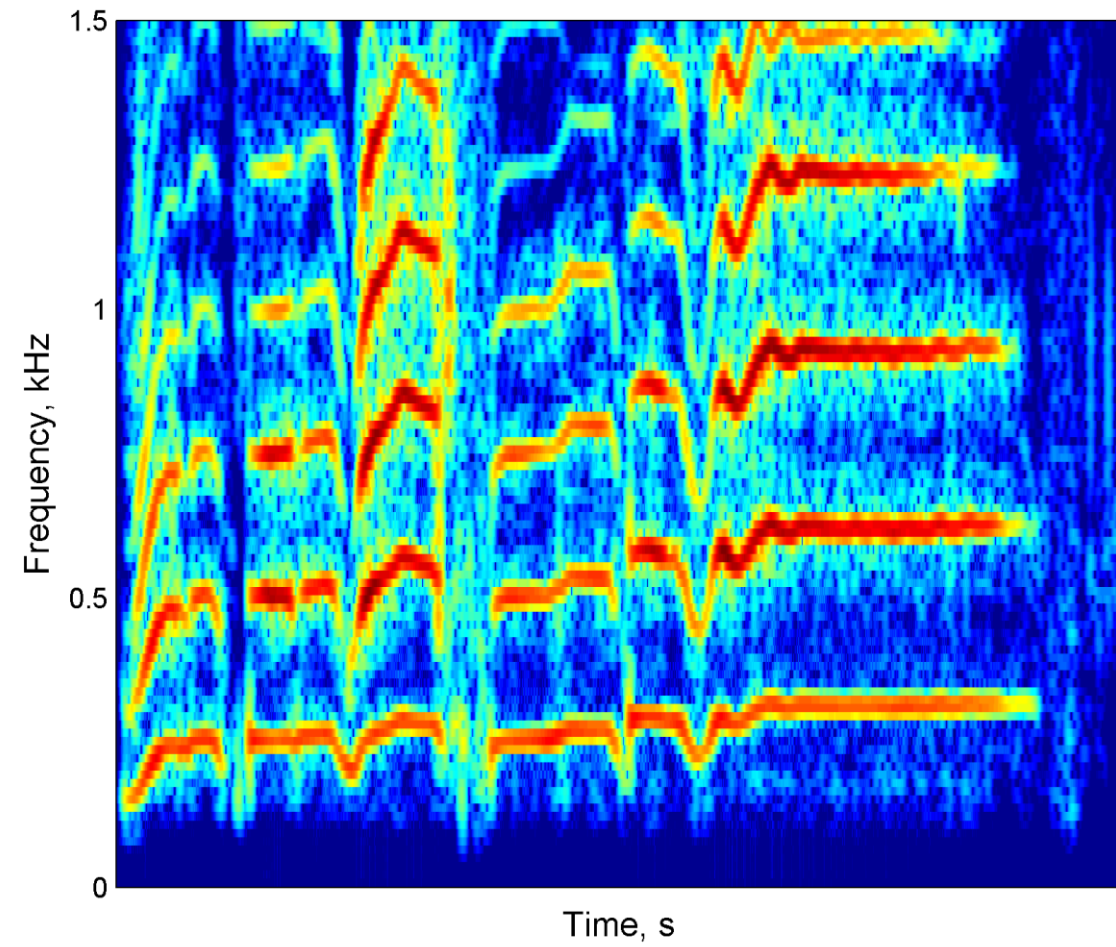
На практике для перерасчета сигнала используется выражение

$$s(t) = \sum_{n=-N_{pt}}^{N_{pt}} s(n) \operatorname{sinc}(t - n).$$



Такой подход 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 работы.

6. Пример обработки сигнала



7. Выводы

- ✓ Рассмотрена задача передискретизации речевого сигнала согласованной с основным тоном. Предложенное решение позволяет выполнять динамическое масштабирование временной оси таким образом, чтобы на каждый период частоты основного тона речевого сигнала приходилось равное число временных отсчетов.
- ✓ Сигнал, получаемый после передискретизации имеет постоянную частоту основного тона. Вследствие этого процедура оценка параметров отдельных гармоник сигнала может быть значительно упрощена.