Автоматическое определение ЧОТ

П. А. Холявин

p.kholyavin@spbu.ru

Автокорреляция

$$r_x(\tau) \equiv \int x(t)\,x(t+\tau)\,dt$$

где x(t) – сигнал, т – задержка (lag)

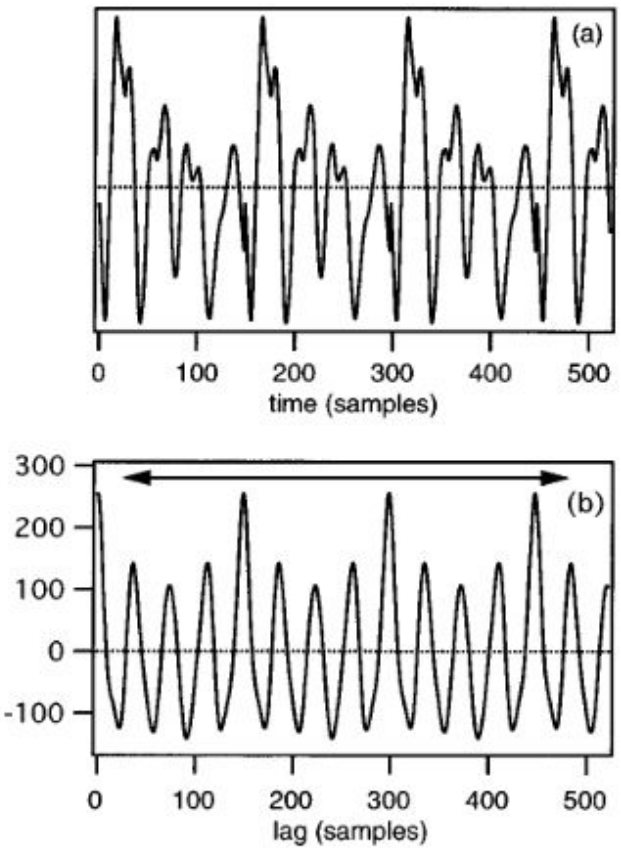
Для дискретного сигнала:

$$r_x[\tau] = \sum_{t=0}^{N-1} x[t] \cdot x[t+\tau]$$

где N – размер окна.

Нормализованная автокорреляция:

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)}$$



Алгоритм Praat

Boersma, Paul. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound." Proceedings of the institute of phonetic sciences. Vol. 17. No. 1193. 1993.

Для каждого окна:

1. Вычитание среднего, умножение на оконную функцию (окно Ханна)

2. Вычисление нормализованной автокорреляции

3. Деление на автокорреляцию самого окна и поиск максимума:

Алгоритм Praat

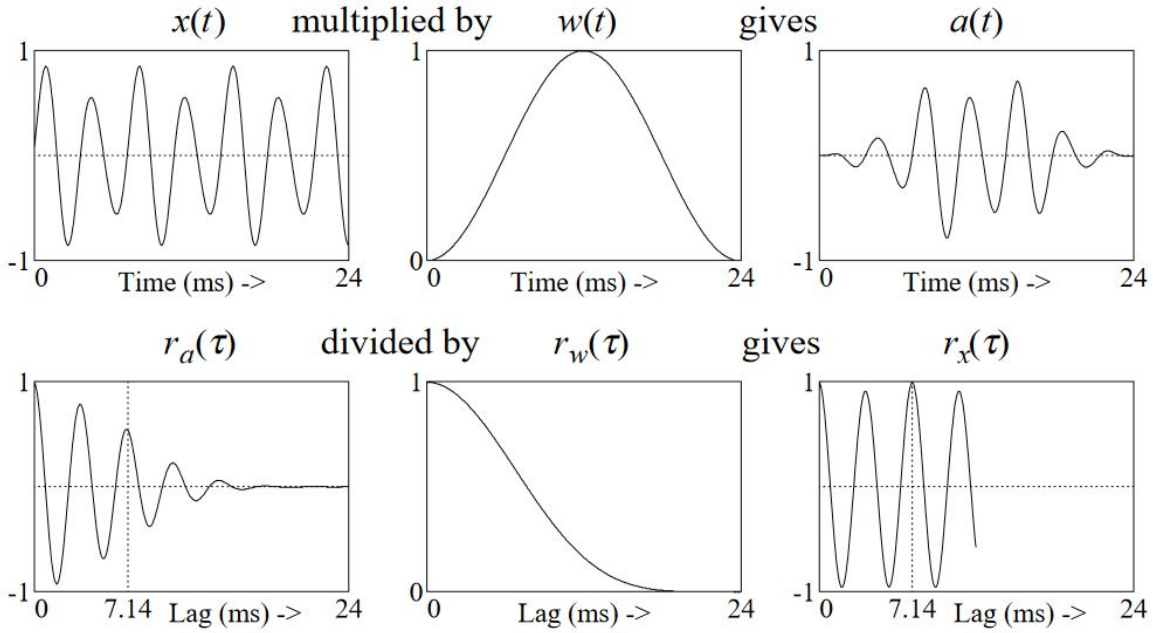


Fig. 1. How to window a sound segment, and how to estimate the autocorrelation of a sound segment from the autocorrelation of its windowed version. The estimated autocorrelation $r_x(\tau)$ is not shown for lags longer than half the window length, because it becomes less reliable there for signals with few periods per window.

Алгоритм Praat

Более подробно:
1. Soft upsampling: FFT → линейное умножение до нуля от 0.95 до 1 частоты Найквиста → IFFT порядка на 1 выше, чем FFT
2. Найти глобальный максимум сигнала
3. Вычисление оконным методом: для каждого фрейма смотрим на <= M “кандидатов” на длину периода, считая “глухой” кандидат
3.1. Извлечение фрейма: длина фрейма = минимальная длина периода * 3
3.2. Вычесть постоянную составляющую
3.3. Первый кандидат – глухой (заданы пороги звонкости и тишины)
3.4. Умножить на оконную функцию
3.5. Добавить ½ фрейма нулей
3.6. Добавить ещё нулей, пока количество отсчётов не станет степенью 2.
3.7. FFT
3.8. Возвести в квадрат

Алгоритм Praat

3.9. IFFT, что даст нам автокорреляцию
3.10. Разделить на АК окна (т.е. пп. 3.5 – 3.9 надо проделать для окна)
3.11. Найти максимумы и их значения, для каждого определить “силу”:
Для глухого:

$$R \equiv VoicingThreshold + \max\left(0, 2 - \frac{(local\ absolute\ peak)/(global\ absolute\ peak)}{SilenceThreshold/(1 + VoicingThreshold)}\right)$$

Для остальных:

$$R \equiv r(\tau_{max}) - OctaveCost \cdot {}^{2}\log(MinimumPitch \cdot \tau_{max})$$

VoicingThreshold = 0.4, SilenceThreshold = 0.05, OctaveCost = 0.01

Алгоритм Praat

4. Т.о. для каждого фрейма n у нас есть p_n кандидатов. Найдём наилучший путь через все фреймы с помощью динамического программирования:

$$cost(\{p_n\}) = \sum_{n=2}^{numberOfFrames} transitionCost\left(F_{n-1,p_{n-1}}, F_{np_n}\right) - \sum_{n=1}^{numberOfFrames} R_{np_n}$$

$$transitionCost(F_1, F_2) = \begin{cases} 0 & \text{if } F_1 = 0 \text{ and } F_2 = 0 \\ VoicedUnvoicedCost & \text{if } F_1 = 0 \text{ xor } F_2 = 0 \\ OctaveJumpCost \cdot \left|{}^2\log \frac{F_1}{F_2}\right| & \text{if } F_1 \neq 0 \text{ and } F_2 \neq 0 \end{cases}$$

Алгоритм REAPER

https://github.com/google/REAPER

1. Вычисление ошибки предсказания LPC и её нормализация

2. Каждый отрицательный пик рассматривается как кандидат на момент закрытия голосовых связок (glottal closure instant, GCI). Они оцениваются на основании их формы (периоды глоттальной волны должны иметь резкий подъём и плавный спуск)

3. Для каждого кандидата вычисляется нормализованная кросс-корреляция

4. Генерируется граф, по которому ищется наилучший путь с помощью динамического программирования (используются дополнительные признаки)

Алгоритм REAPER

Признаки:

1. Псевдовероятность звонкости (энергия на низких частотах)
2. Псевдовероятность начала и конца звонкого участка (изменение энергии)

Алгоритм REAPER

Динамическое программирование

```
For each pulse in the utterance:
  For each period hypotheses following the pulse:
    For each period hypothesis preceding the pulse:
      Score the transition cost of connecting the periods.  Choose the
      minimum overall cost (cumulative+local+transition) preceding
      period hypothesis, and save its cost and a backpointer to it.
      The costs of making a voicing state change are modulated by the
      probability of voicing onset and offset.  The cost of
      voiced-to-voiced transition is based on the delta F0 that
      occurs, and the cost of staying in the unvoiced state is a
      constant system parameter.
```

Спасибо за внимание!