

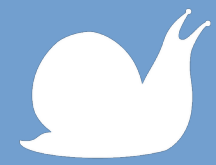
# Моделирование мелодического контура

П. А. Холявин

p.kholyavin@spbu.ru

31.10.2024





# Автоматическая интонационная разметка

1. По правилам
2. С помощью машинного обучения



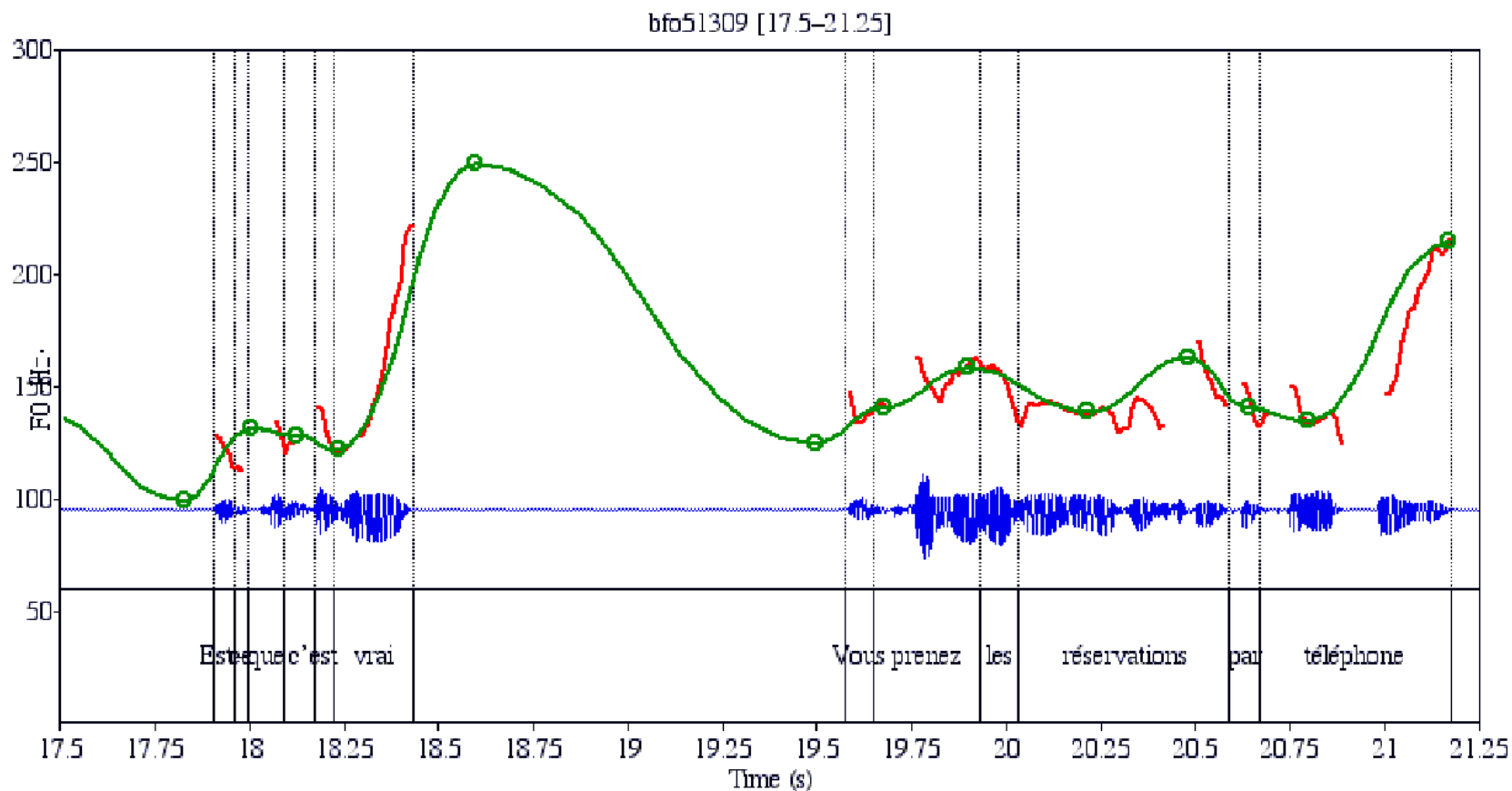
# Алгоритм Momel-INTSINT

## Momel

0. Мелодический контур можно разделить на микропросодический и макропросодический компоненты
1. Макропросодический компонент моделируется как квадратичный сплайн
2. Сплайн определяется последовательностью точек  $\langle t, h, k \rangle$ , где  $t$  – время точки,  $h$  – значение ЧОТ в ней,  $k$  – место перегиба (для простоты считается, что место перегиба ровно посередине между двумя точками)



# Алгоритм Momet-INTSINT





# Алгоритм Momet-INTSINT

## Собственно алгоритм

### 1. Предобработка значений ЧОТ

1.1. Значения ЧОТ берутся из контура с шагом 10 мс, в глухих участках значение принимается равным 0

1.2. Если значение более чем на 5 % выше обоих соседей, оно принимается равным 0



# Алгоритм Momet-INTSINT

## 2. Поиск кандидатов на целевые точки

Для каждой точки:

- 2.1. задаётся окно с центром в этой точке (300 мс)
- 2.2. если в окно попали значения меньше или больше заранее заданных пороговых, они убираются
- 2.3. к оставшимся значениям применяется квадратичная регрессия
- 2.4. убираются все точки, значение которых более чем на 5 % ниже, чем предсказано регрессией
- 2.5. пп. 2.3 и 2.4 применяются, пока точки не перестают убираться



# Алгоритм Momet-INTSINT

2.6. с помощью регрессии вычисляется целевая точка:

$$y = ax^2 + bx + c$$

$$t = \frac{-b}{2a}; h = at^2 + bt + c$$

2.7. Если  $t$  вышло за пределы окна или  $h$  вышло за пороговые значения ЧОТ, они не учитываются

Пп. 2.1-2.7 повторяются для каждой точки ЧОТ



# Алгоритм Momel-INTSINT

## 3. Группировка кандидатов в сегменты

Для каждой точки:

3.1. задаётся окно с центром в этой точке (200 мс)

3.2. считаются параметры  $dt(x)$  и  $dh(x)$ :

$dt(x)$  – усреднённое расстояние (*средняя абсолютная разница*) между всеми  $t$  в первой и второй половине окна

$dh(x)$  – то же для  $h$

$$wd = \frac{1}{\text{mean}(dt(x))}; wh = \frac{1}{\text{mean}(dh(x))}$$

$$d(x) = \frac{dt(x) \cdot wd + dh(x) \cdot wh}{wd + wh}$$





# Алгоритм MomeI-INTSINT

3.3. Задаются границы сегментов с условиями:

$$d(x) > d(x - 1)$$

$$d(x) > d(x + 1)$$

$$d(x) > \text{mean}(d(x))$$

Если точка  $x$  удовлетворяет условиям, она считается границей сегмента



# Алгоритм Momet-INTSINT

## 4. Устранение кандидатов

Внутри каждого сегмента:

4.1. Вычисляется среднее  $dt(x)$  и  $dh(x)$

4.2. Каждая точка, у которой один из этих параметров отклоняется от среднего больше, чем на одно стандартное отклонение, устраняется

4.3. После этого вычисляются средние  $t$  и  $h$  у оставшихся точек

4.4. Они и будут параметрами искомой целевой точки



# Алгоритм Momet-INTSINT

## INTSINT

Задаются два параметра: key (“тональность”) в Гц и range (диапазон) в октавах

The targets **T**, **M** and **B** are defined 'absolutely' without regard to the preceding targets

$$\mathbf{T} = \text{key} * \sqrt{2^{\text{range}}}$$

$$\mathbf{M} = \text{key}$$

$$\mathbf{B} = \text{key} / \sqrt{2^{\text{range}}}$$

Other targets are defined with respect to the preceding target:

$$\mathbf{H} = \sqrt{P_{i-1} * \mathbf{T}}$$

$$\mathbf{U} = \sqrt{P_{i-1} * \sqrt{P_{i-1} * \mathbf{T}}}$$

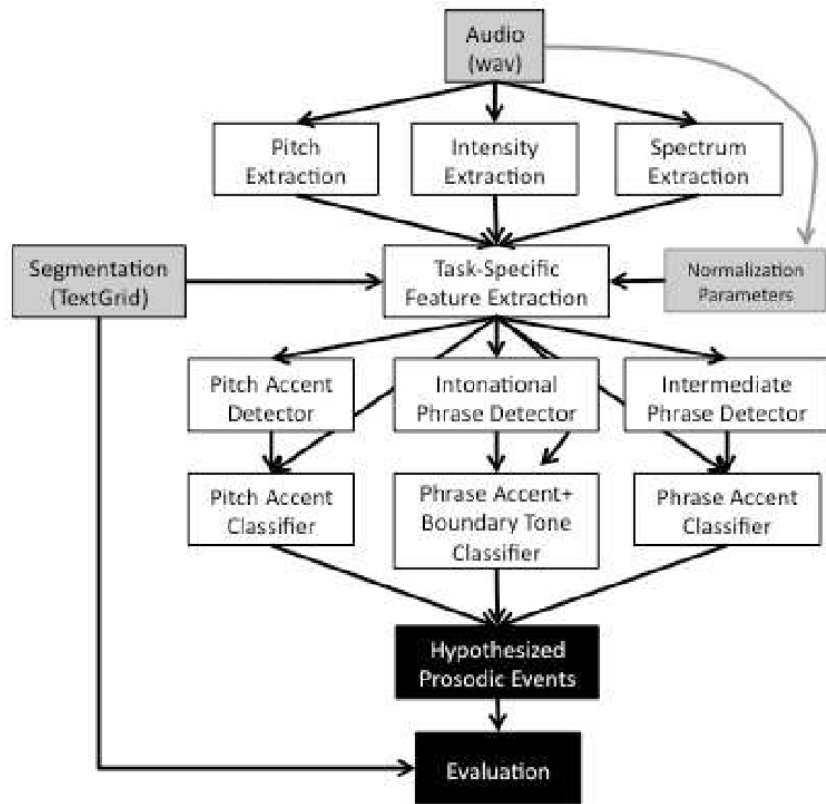
$$\mathbf{S} = P_{i-1}$$

$$\mathbf{D} = \sqrt{P_{i-1} * \sqrt{P_{i-1} * \mathbf{T}}}$$

$$\mathbf{L} = \sqrt{P_{i-1} * \mathbf{B}}$$



# Алгоритм AuToVi





# Алгоритм AuToVI

## 1. Детекция Pitch Accent

Классификатор: логистическая регрессия

Признаки: среднее, минимум, максимум, стандартное отклонение, z-score максимума для ЧОТ, нормализованной ЧОТ по диктору, интенсивности и их производных

энергия на частоте от 2 до 20 барк, отношение её к общей энергии в окне

Окно: слово и N его соседей (0, 1, 2) справа и слева



# Алгоритм AuToVI

## 2. Классификация Pitch Accent

Классификатор: метод опорных векторов

Признаки: среднее, минимум, максимум, стандартное отклонение, z-score максимума для ЧОТ, нормализованной ЧОТ по диктору, интенсивности и их производных

Окно: самый громкий псевдо-слог в анализируемом слове



# Алгоритм AuToVI

## 3. Определение границ фраз

Классификатор: AdaBoost, метод опорных векторов

Признаки: среднее, минимум, максимум, стандартное отклонение, z-score максимума для ЧОТ, нормализованной ЧОТ по диктору, интенсивности и их производных

разница в признаке между текущим словом и следующим  
паузы

Окно: самый громкий псевдо-слог в анализируемом слове



# Алгоритм AuToVI

## 4. Классификация границ фраз

Классификатор: метод опорных векторов

Признаки: среднее, минимум, максимум, стандартное отклонение, z-score максимума для ЧОТ, нормализованной ЧОТ по диктору, интенсивности и их производных

Окно: последние 200 мс последнего слова перед границей



Спасибо за внимание!

