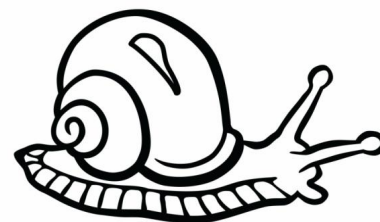


# Автоматический синтез речи. Компилятивный синтез

П. А. Холявин

p.kholyavin@spbu.ru

??..04.2024





# Типы единиц компилятивного синтеза

1. Аллофоны
2. Дифоны
3. Субаллофоны
4. Слоги
5. Слова... (макросинтез)
6. Unit Selection (селективный синтез)



# Аллофонный синтез

Согласные:

L \ R	Неогубл. гласный	Огубл. гласный	Глухой согл.	Сонант или /v/	Звонкий согл.
Абс. начало					
Гласный					
Согл.					

Сочетания согласных: /b/, /p/, /m/, /d/, /t/ + /n/, /n'/



# Аллофонный синтез

Гласные – левые контексты:

/f/, /v/, /p/, /b/	/n'/
/f'/, /v'/, /p'/, /b'/	/l/
/t/, /d/, /s/, /z/, /ʃs/	/l'/
/t'/, /d'/, /s'/, /z'/, /tʃ/, /ʃ/, /r'/	/r/
/ʃ/, /ʒ/	/j/
/k/, /g/, /x/	/a/
/k'/, /g'/, /x'/	/o/
Абс. начало	/u/
/m/	/e/
/m'/	/i/
/n/	/i'/



# Аллофонный синтез

Гласные – правые контексты:

/p/, /b/, /m/, /f/, /v/		/j/
/t/, /d/, /s/, /z/, /ʃs/, /n/, /r/		/a/
/ʃ/, /ʒ/		/o/
/k/, /g/, /x/		/u/
Мягкие, кроме /j/		/e/
Абсолютный конец		/i/
//		/ɨ/



# Аллофонный синтез

Гласные – степени редукции:

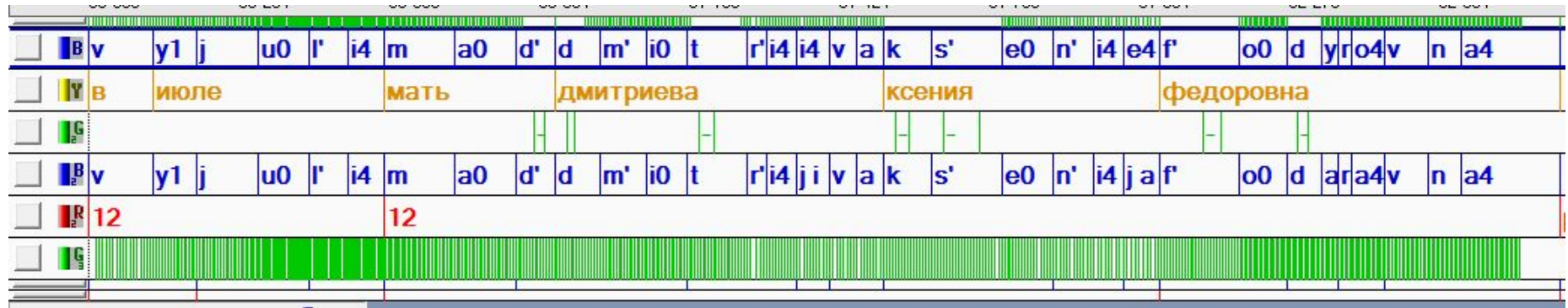
0	Ударный гласный
1	Предударные гласные
2	/a/ во втором предударном слоге
4	Заударные гласные
5	Заударный неконечный /a/

<i>вертолёт</i>	[v'i <sub>1</sub> rt <sub>1</sub> l'o <sub>0</sub> t]
<i>барабан</i>	[ba <sub>2</sub> ra <sub>1</sub> ba <sub>0</sub> n]
<i>подушка</i>	[pa <sub>1</sub> du <sub>0</sub> ʃka <sub>4</sub> ]
<i>подушках</i>	[pa <sub>1</sub> du <sub>0</sub> ʃka <sub>5</sub> x]



# Корпус CORPRES

Skrelin, Pavel, et al. "CORPRES" *International Conference on Text, Speech and Dialogue*. Springer, Berlin, Heidelberg, 2010.  
(CORpus of Professionally REad Speech)





# Другие методы синтеза

**Дифонный:** вторая половина звука А + первая половина звука Б

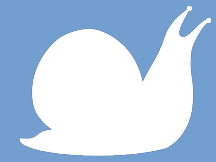
+ лёгкость сегментации

- увеличение числа единиц

**Субаллофонный:** половины звуков

**Слоговой:** слоги (количество единиц зависит от языка)





# Unit Selection

Метод синтеза речи по тексту, в котором вся база данных потенциально может быть использована в качестве единиц синтеза.

Метод Unit Selection основан на **принципе наименьшей модификации**.

Используются два параметра:

1. Стоимость разницы  $T$  (насколько потенциальный сегмент  $u$  отличается от целевого  $s$ )
2. Стоимость склейки  $J$  (насколько плохо соседние сегменты сочетаются друг с другом)

$$C(U, S) = \sum_{t=1}^T T(u_t, s_t) + \sum_{t=1}^{T-1} J(u_t, u_{t+1})$$



# Unit Selection

Единицы Unit Selection:

окна, субаллофоны, дифоны, аллофоны, полуслоги, дислоги, слоги, слова, фразы,  
...

Возможны системы, где используются несколько типов единиц.

Спасибо за внимание!

