

Автоматический синтез речи. Статистический параметрический синтез

П. А. Холявин

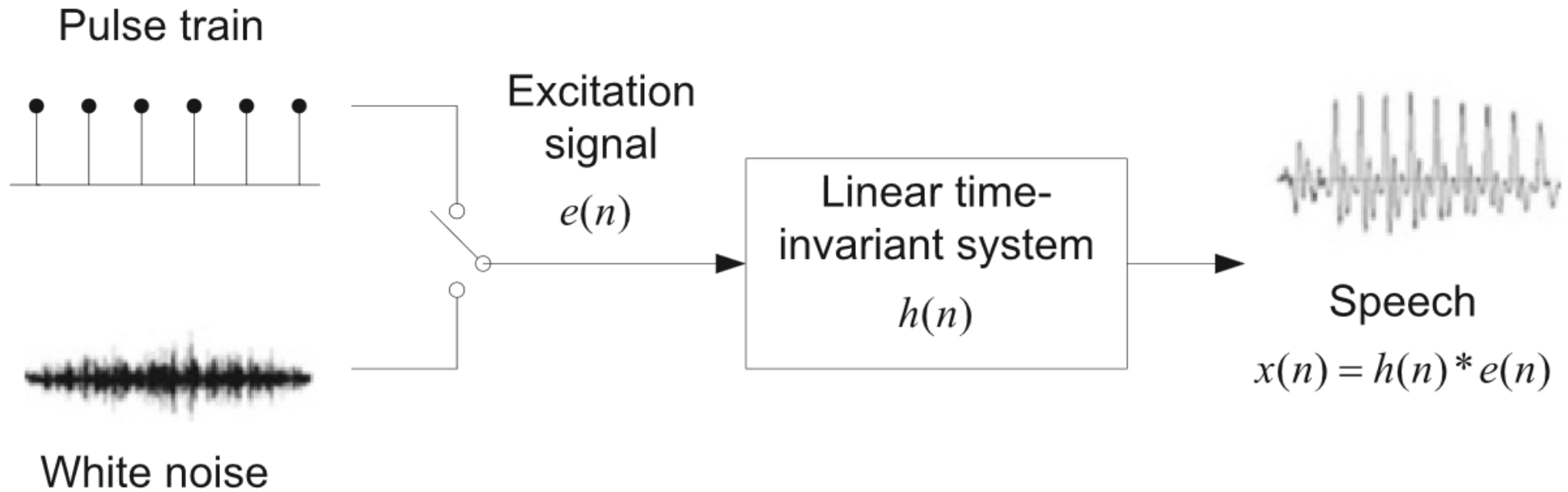
p.kholyavin@spbu.ru

01.04.2024



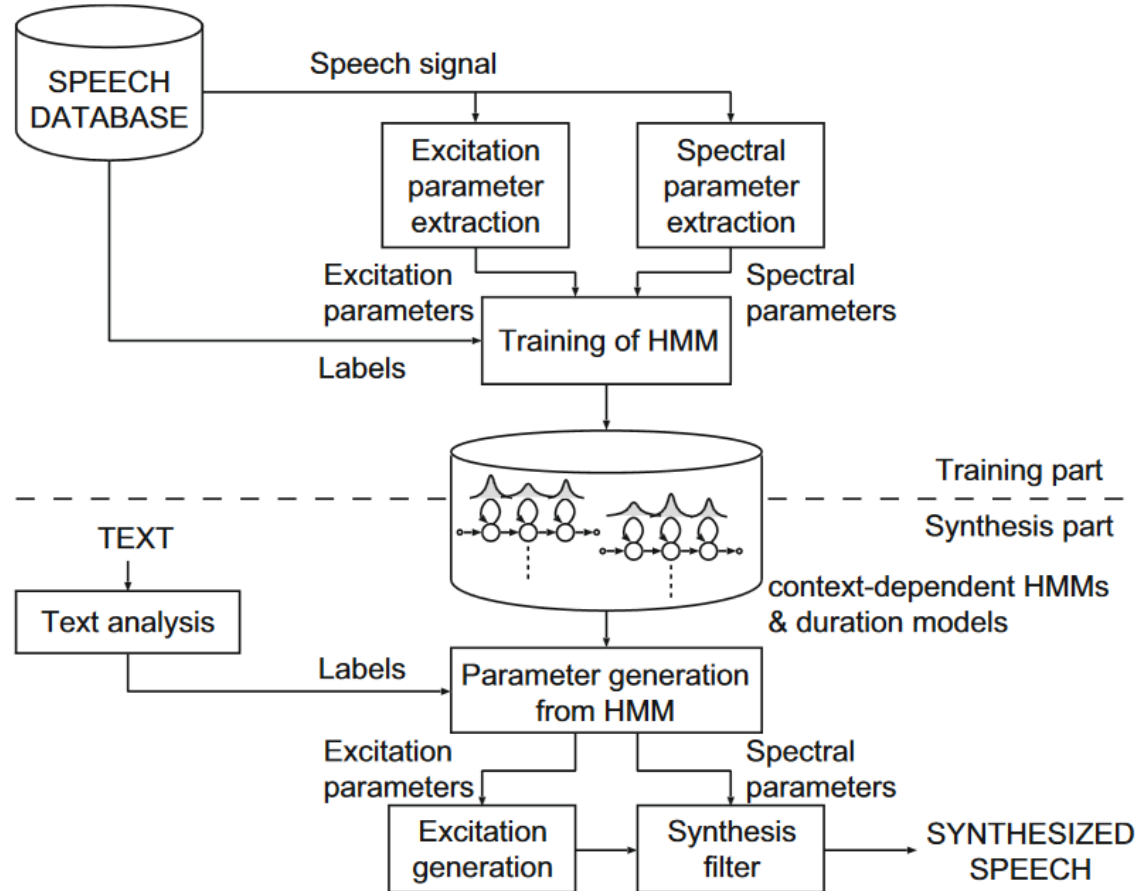


Модель источник-фильтр



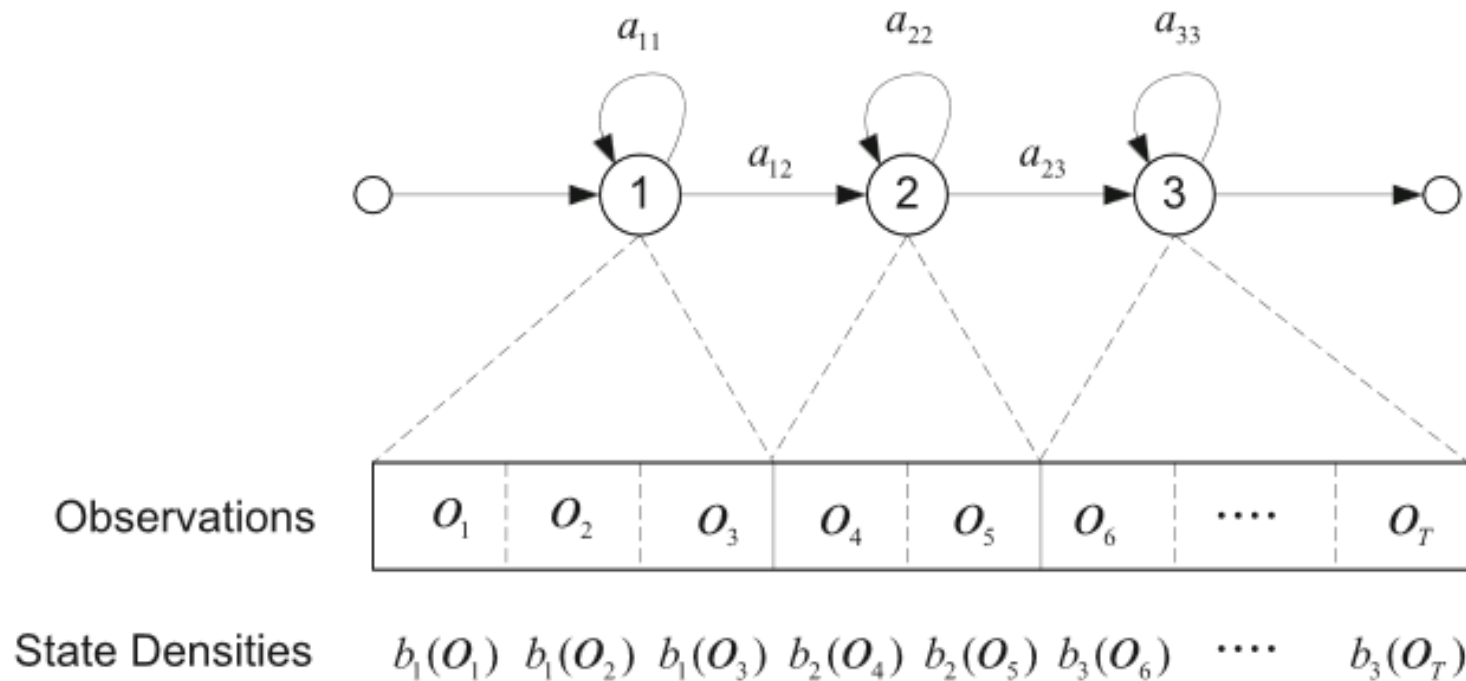


Скрытые Марковские модели





Скрытые Марковские модели





Скрытые Марковские модели

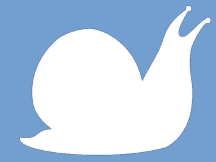
Формально:

Обучение:

$$\hat{\lambda} = \arg \max_{\lambda} \{p(\mathbf{O}|\mathcal{W}, \lambda)\}$$

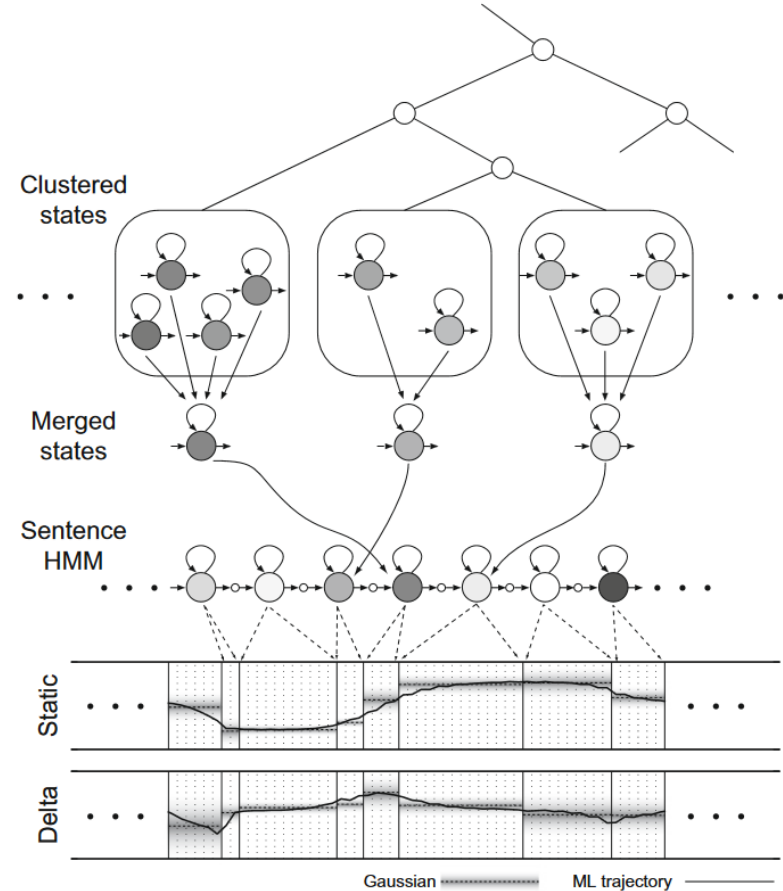
Синтез:

$$\hat{o} = \arg \max_o \left\{ p(o|w, \hat{\lambda}) \right\}$$



Скрытые Марковские модели

Проблема: генерируемый сигнал
должен быть гладким





Линейное предсказание

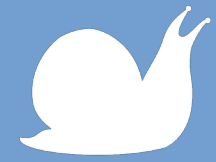
$$y[n] = x[n] \otimes h[n]$$

$$Y(z) = X(z)H(z)$$

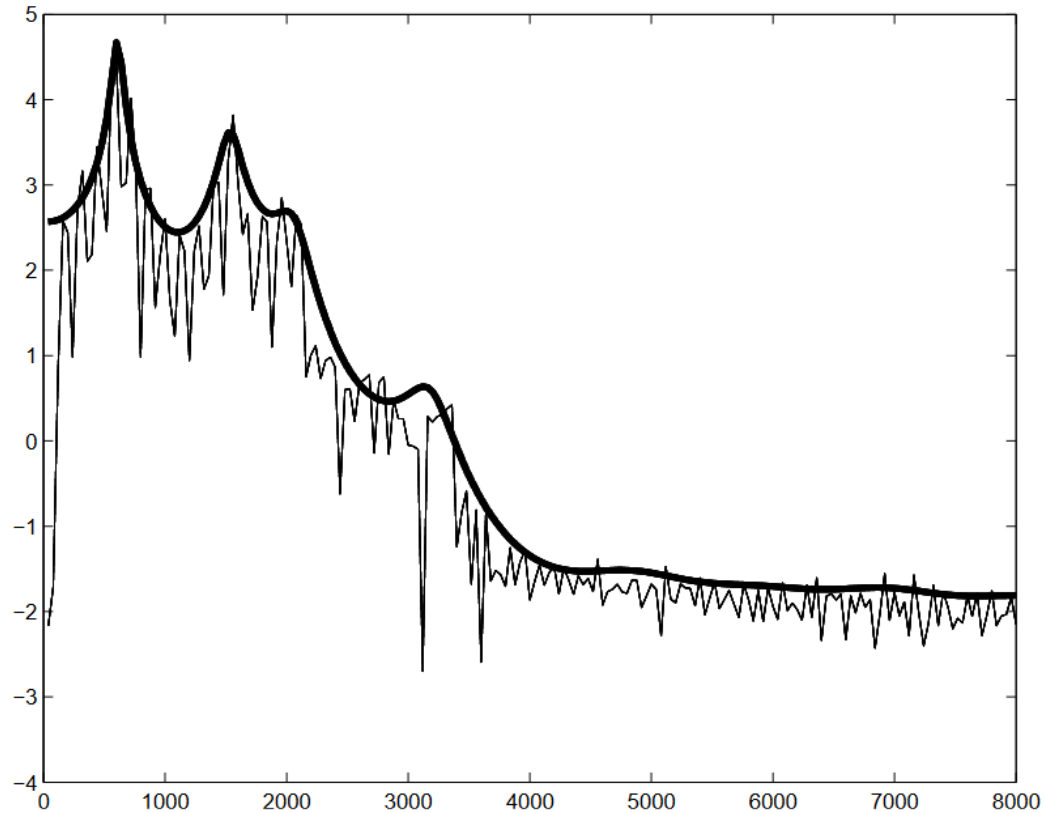
$$= \frac{1}{1 - \sum_{k=1}^P a_k z^{-k}} X(z)$$

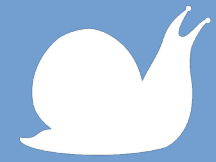
$$y[n] = \sum_{k=1}^P a_k y[n-k] + x[n]$$

$$\tilde{y}[n] = \sum_{k=1}^P a_k y[n-k]$$



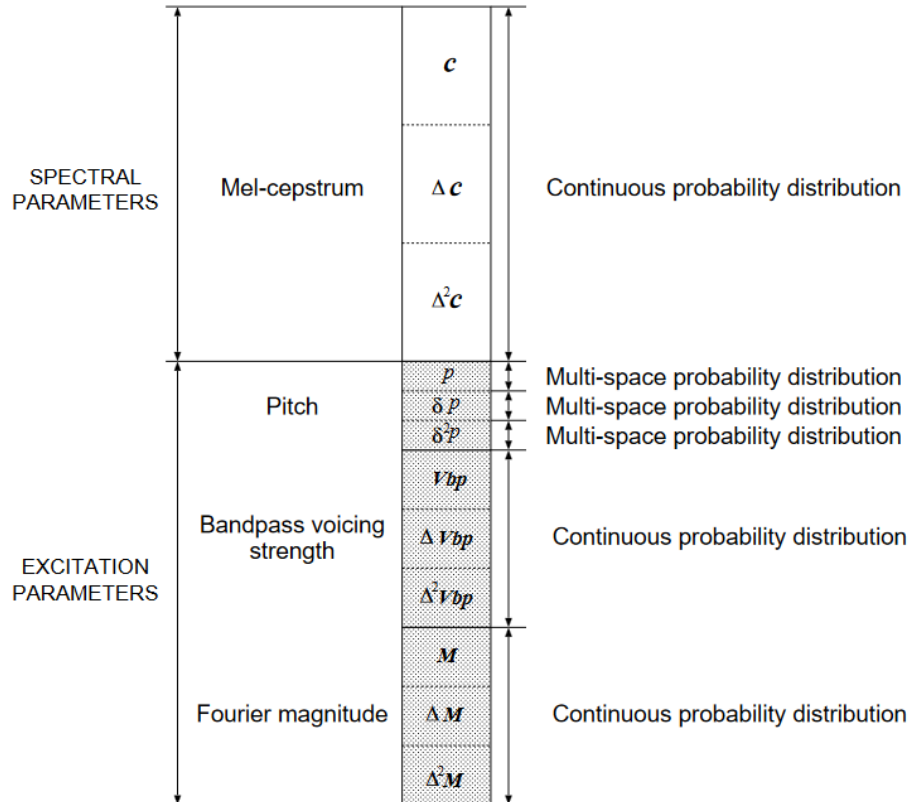
Линейное предсказание





Моделирование источника

1. Mixed excitation

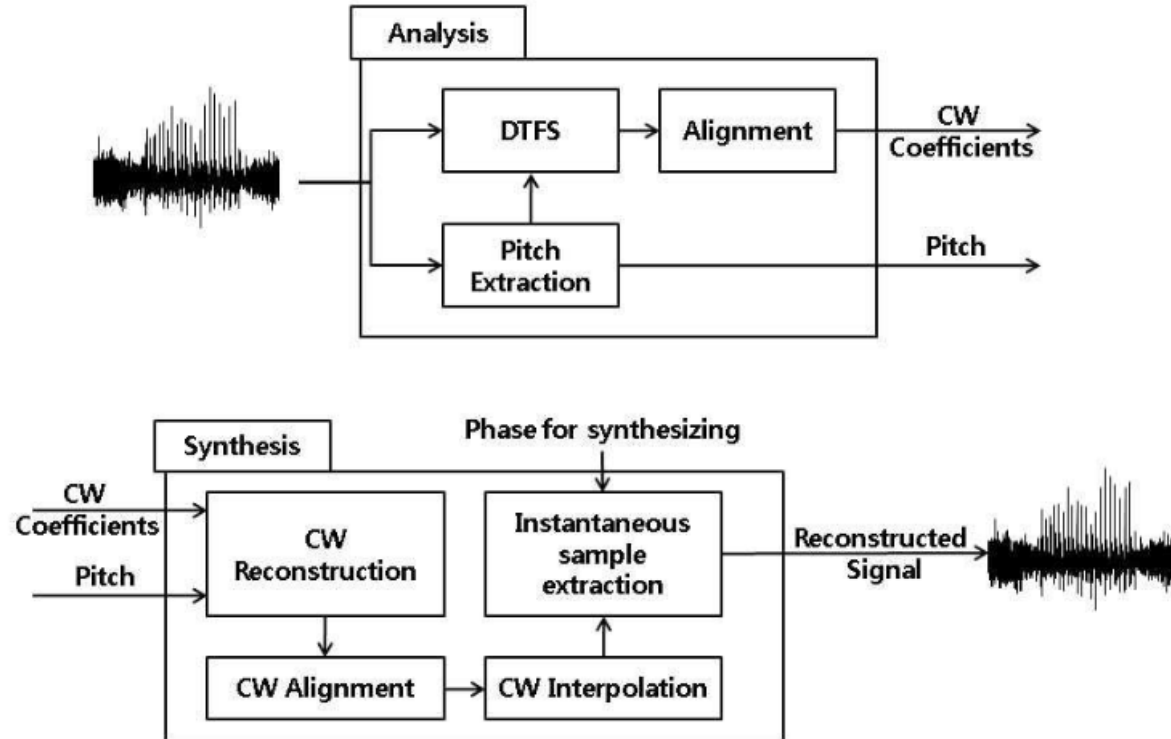




Моделирование источника

2. Waveform interpolation:

CW — characteristic waveform





Моделирование источника

3. Liljencrants-Fant (LF) model

$$e_{LF}(t) = \begin{cases} E_0 e^{\alpha t} \sin(w_g t), & t_o \leq t \leq t_e \\ -\frac{E_e}{\epsilon T_a} [e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}], & t_e < t \leq t_c \\ 0, & t_c < t \leq T_0 \end{cases}$$

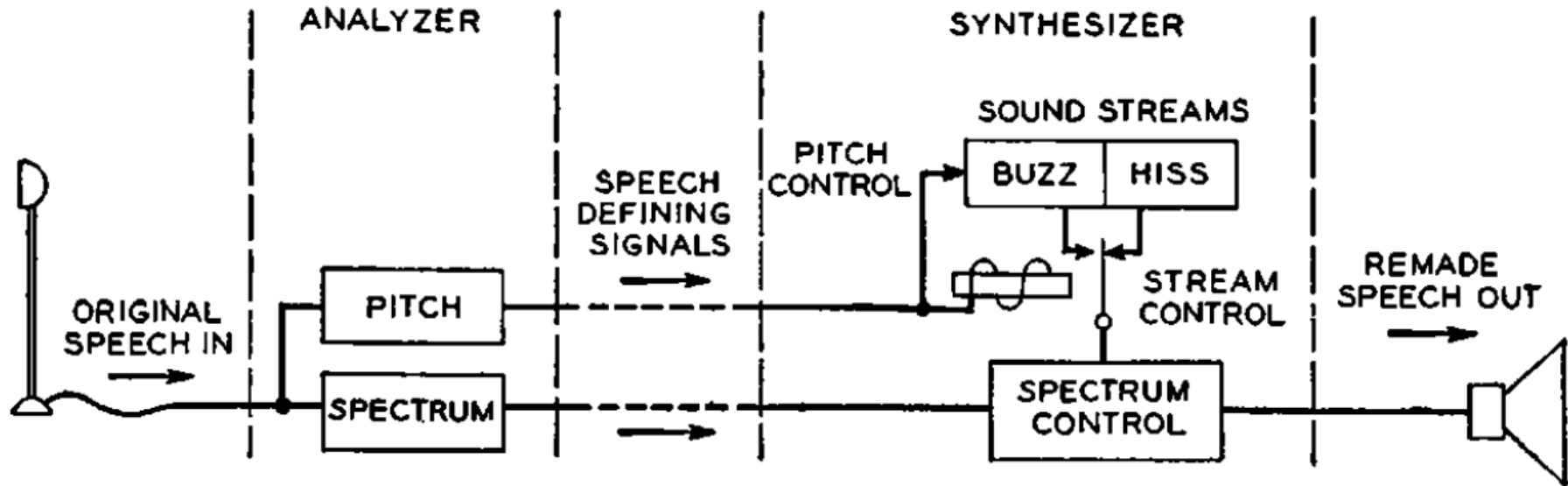


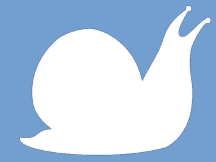
Моделирование фильтра

1. Мел-кепстральные коэффициенты (MFCC)
2. Мел-спектрограммы
3. Барк-кепстральные коэффициенты
4. Band aperiodicity

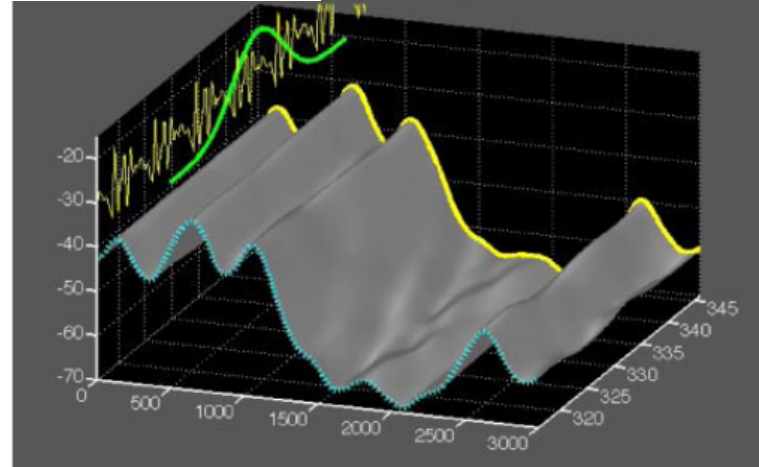
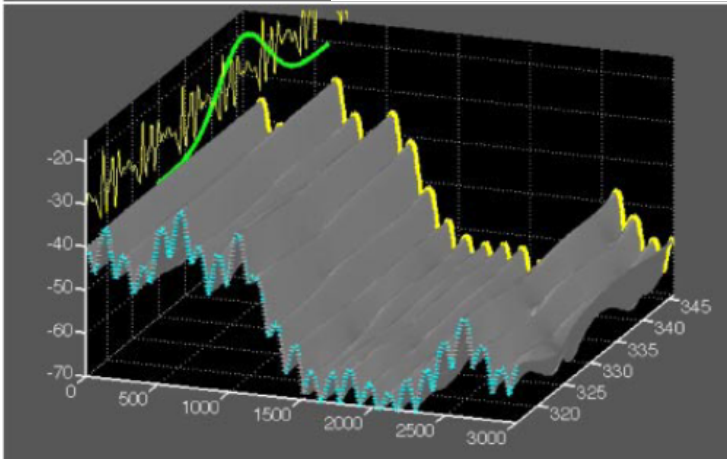
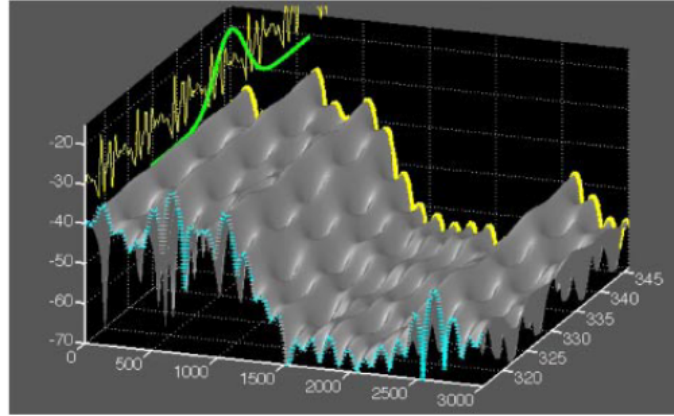


Вокодеры: VOCODER



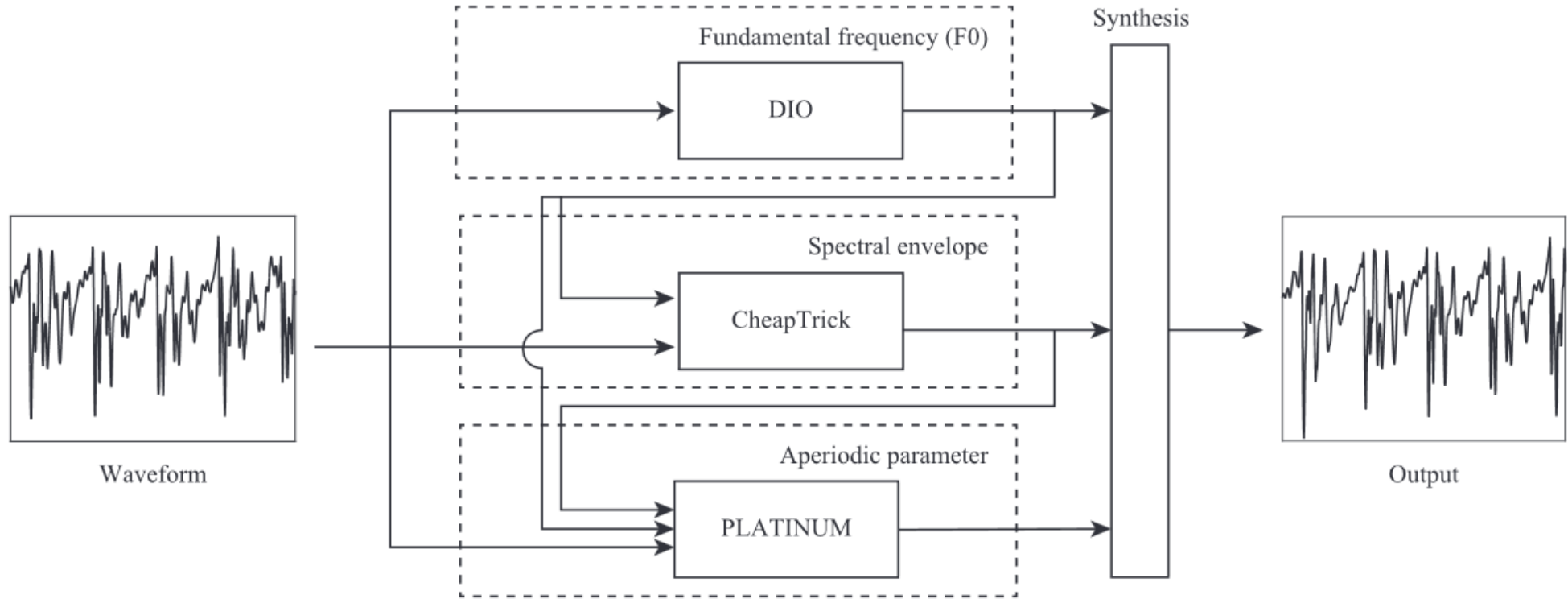


Вокодеры: STRAIGHT



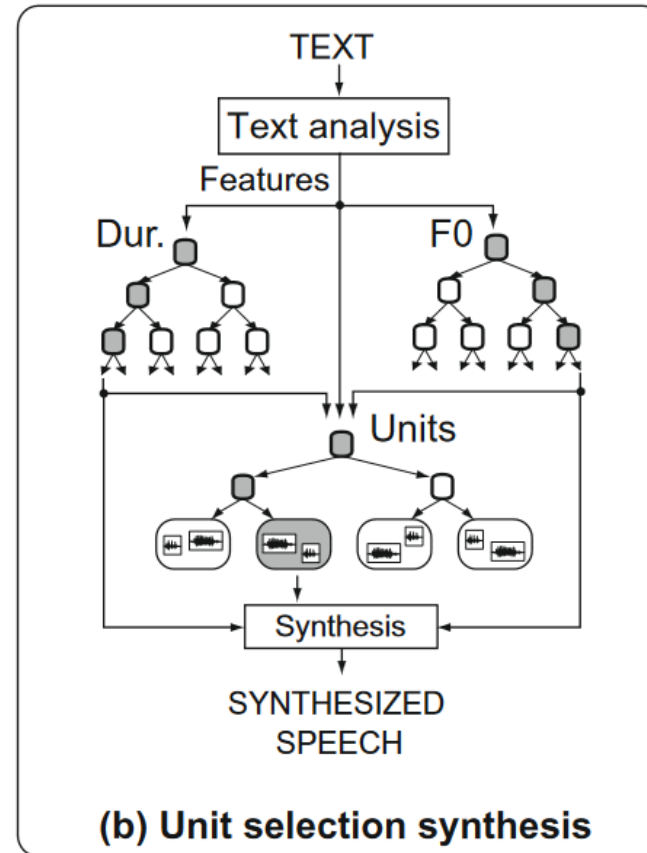
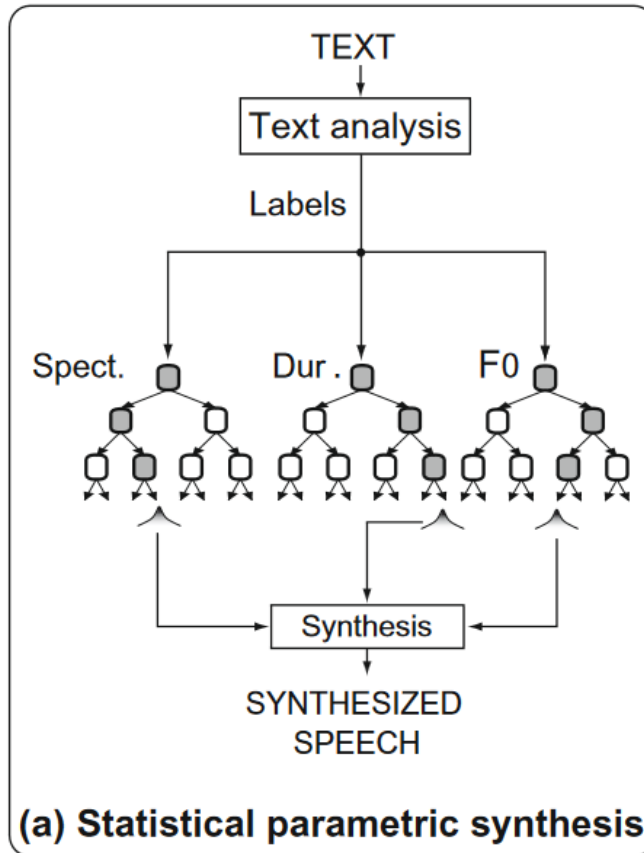


Вокодеры: WORLD





Деревья принятия решений



Спасибо за внимание!

