РАЗДЕЛЬНОЕ МОДЕЛИРОВАНИЕ РЕЧЕВОГО СООБЩЕНИЯ В ВИДЕ ГОЛОСОВЫХ, ФОНЕТИЧЕСКИХ И ПРОСОДИЧЕСКИХ ПАРАМЕТРОВ

И.С. Азаров, А.А. Петровский

Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

OSTIS-2015

1. Моделирование речевого сигнала

Основные приложения:

- распознавание речи;
- верификация диктора;
- кодирование;
- очистка от шума;
- повышение разборчивости;
- синтез речи по тексту;
- конверсия голоса;

Уровни моделирования

- Низкоуровневое моделирование сигнала (параметры гармоник, СПМ);
- Моделирование специфических речевых характеристик (речевой тракт);
- Моделирование высокоуровневых речевых характеристик (голос, акцент, экспрессия, смысловое и фонетическое содержание).

2. Выделение информации из речевого сигнала

Сложность автоматического выделения информации обусловлена нестационарностью (изменчивостью параметров):

На уровне сигнала

- изменение интенсивности и периодичности;
- изменение частоты основного тона, параметров гармоник, СПМ;

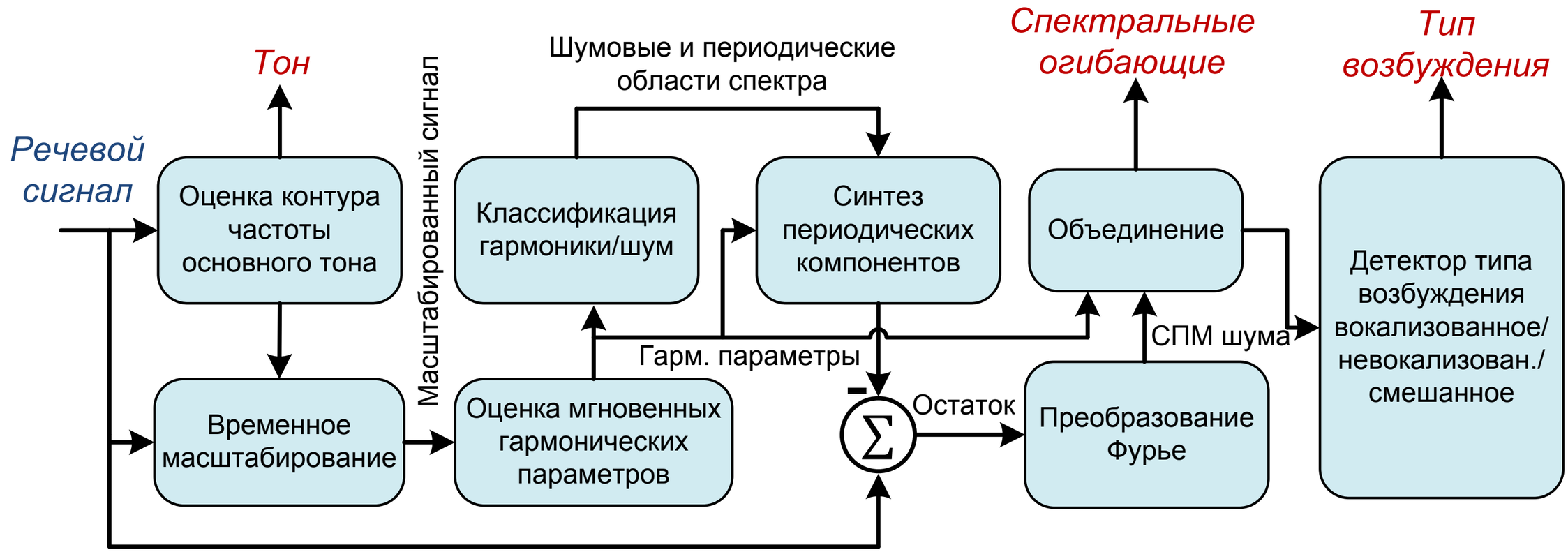
На уровне речевого тракта

- изменение источника сигнала (вокализованные/невокализованные звуки);
- изменение режимов и параметров фонации;

На уровне диктора

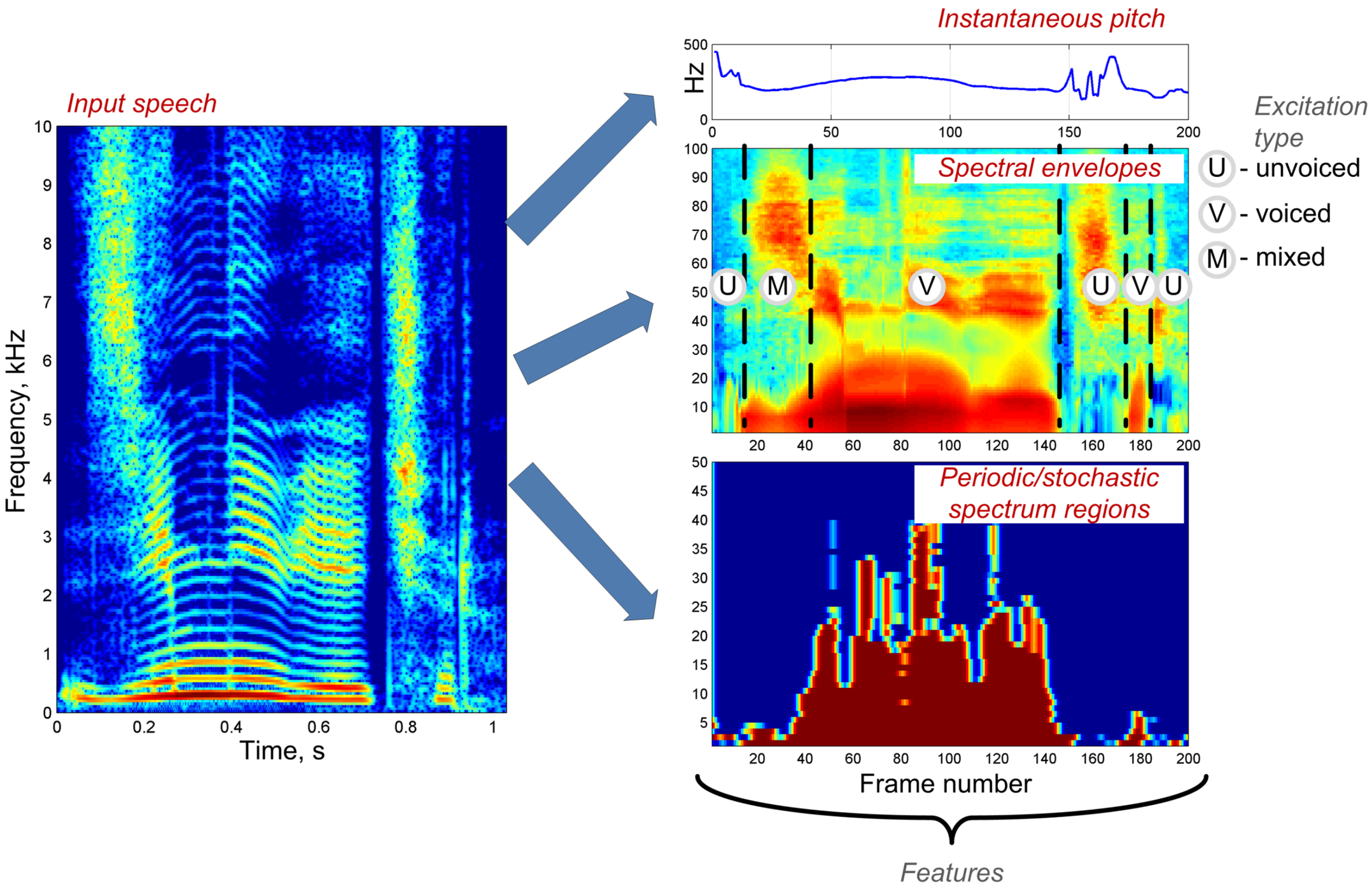
- голос: тембр голоса, интонация;
- особенности произношения, акцент;
- темп речи;

3. Моделирование сигнала

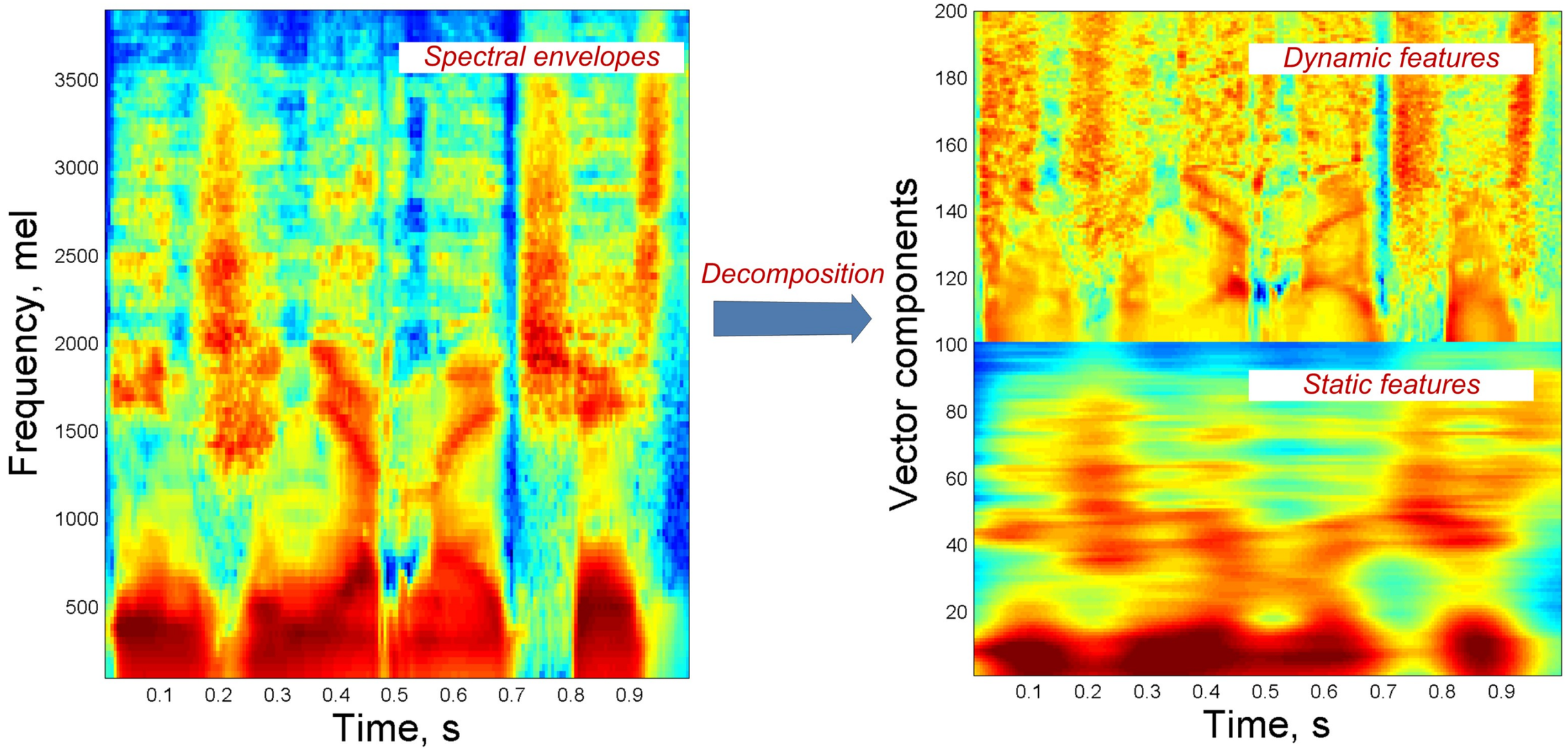


Алгоритм: 1) оценка ОТ; 2) временное масштабирование; 3) оценка гармонических параметров; 4) оценка периодичности; 5) синтез периодической части; 6) оценка СПМ шума; 7) объединение огибающей; 8) определение типа возбуждения.

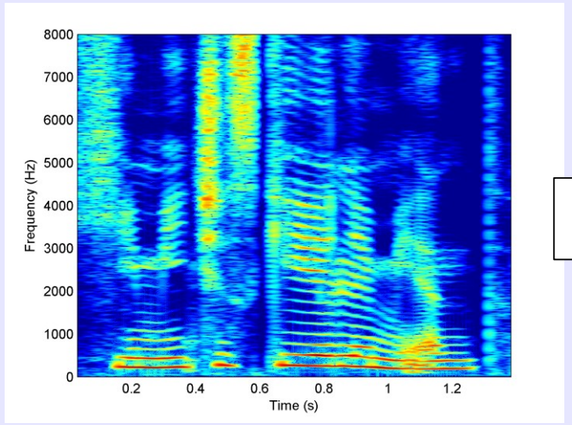
4. Параметрическое описание сигнала

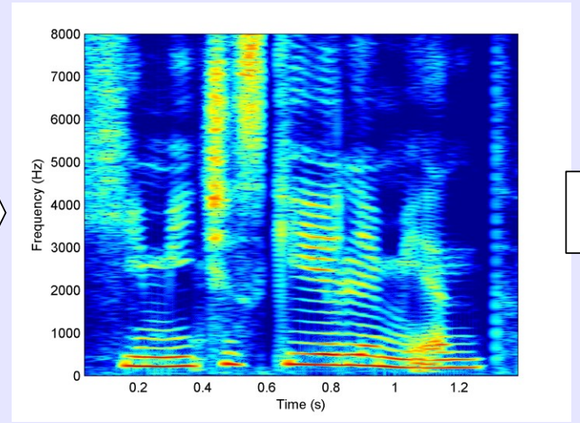


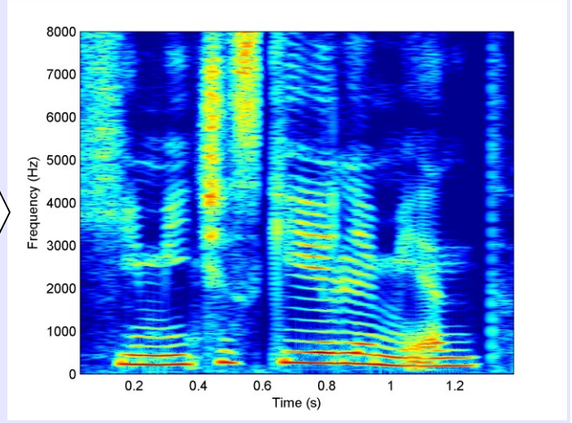
5. Огибающие амплитудного спектра

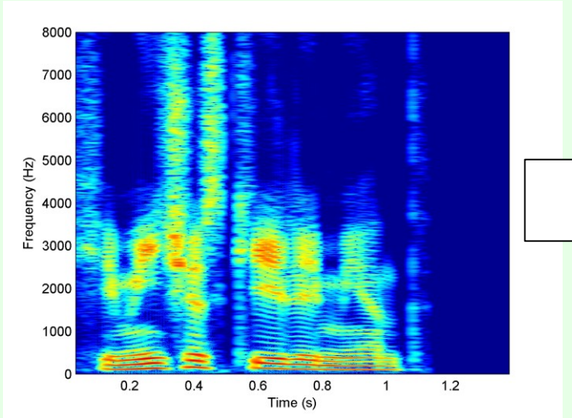


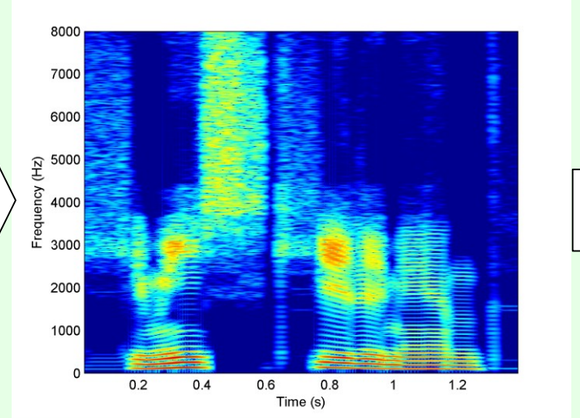
Для повышения информативности характеристических векторов выполняется разделение на низкочастотные и высокочастотные компоненты (используется фильтр с частотой среза 4–9Гц). НЧ компоненты больше связаны с голосовыми признаками, а ВЧ с фонетическими и артикуляторными.

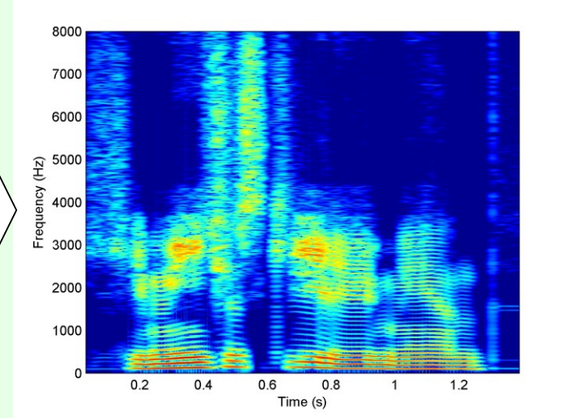












6. Выравнивание темпа

Итерация 1

Итерация 5

Исходный
(женский
голос)

Целевой
(мужской
голос)

7. Основная идея

Идея

Создание параметрического описания речевого сигнала, основанного на раздельном моделировании голоса диктора и содержания речевого сообщения.

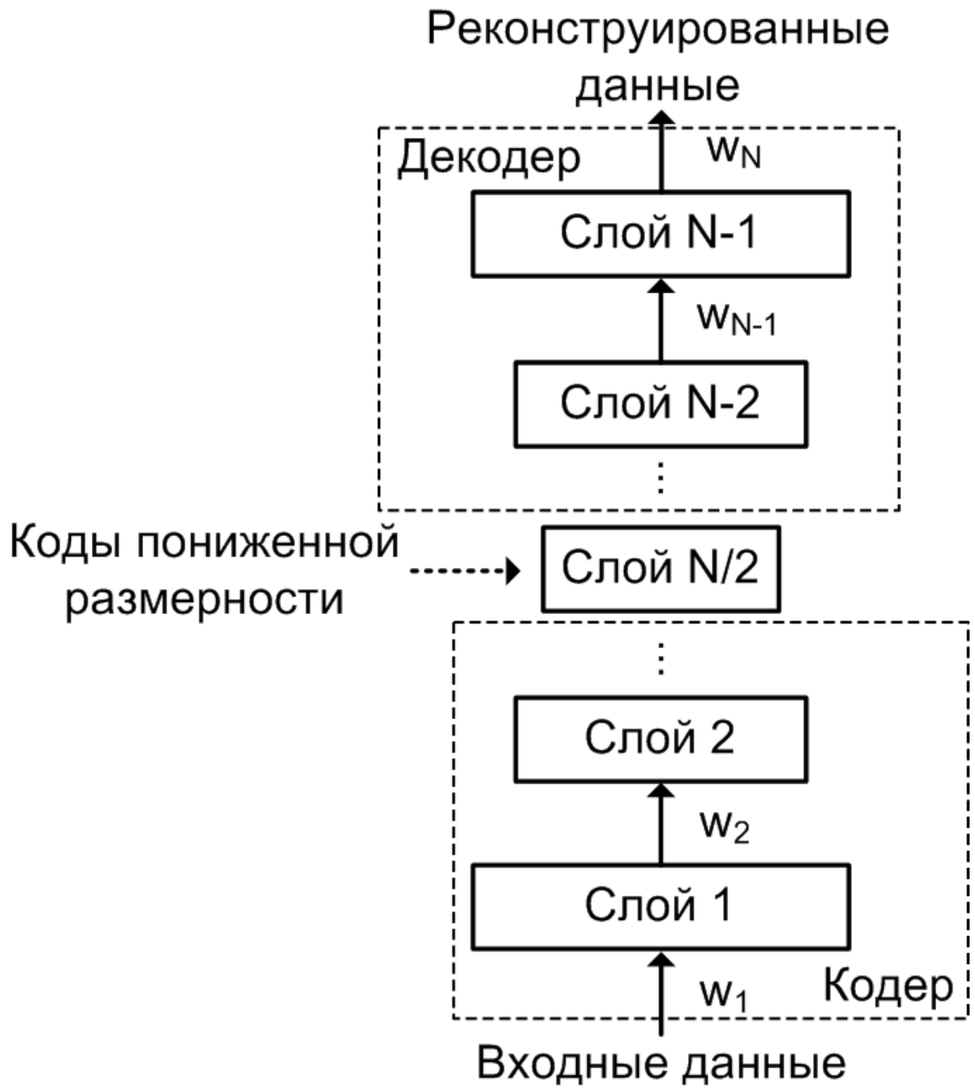
Модель

Вместо преобразований речь->текст и текст->речь использовать преобразование речь->фонемы+просодика и речь->фонемы+просодика

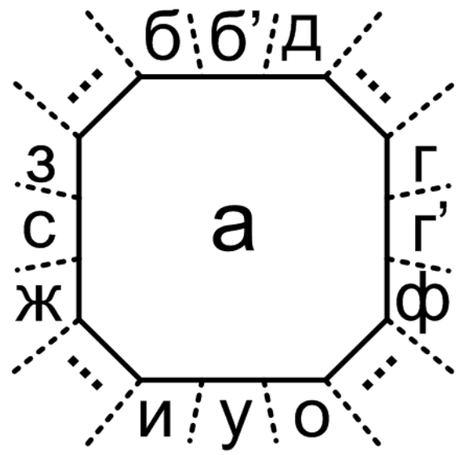
Реализация

Фиксирование фонем и автоматическое упорядочивание характеристических векторов.

8. Использование автоматического кодера

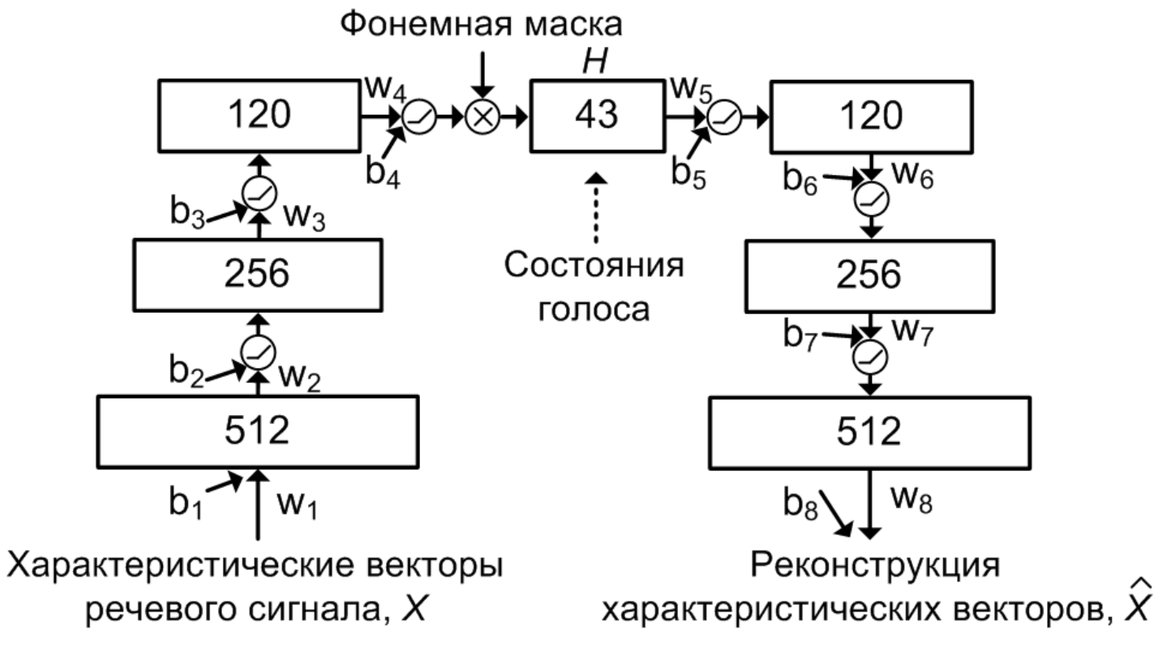


Автоматический кодер



Пространство кодов пониженной размерности

9. Конфигурация нейронной сети



Конфигурация нейронной сети

Фонемная маска кодов

$$H = (w_4 RL(w_3 RL(w_2 RL(w_1 X + b_1) + b_2) + b_3) + b_4) \otimes M$$

$$\hat{X} = (w_8 RL(w_7 RL(w_6 RL(w_5 H + b_5) + b_6) + b_7) + b_8)$$

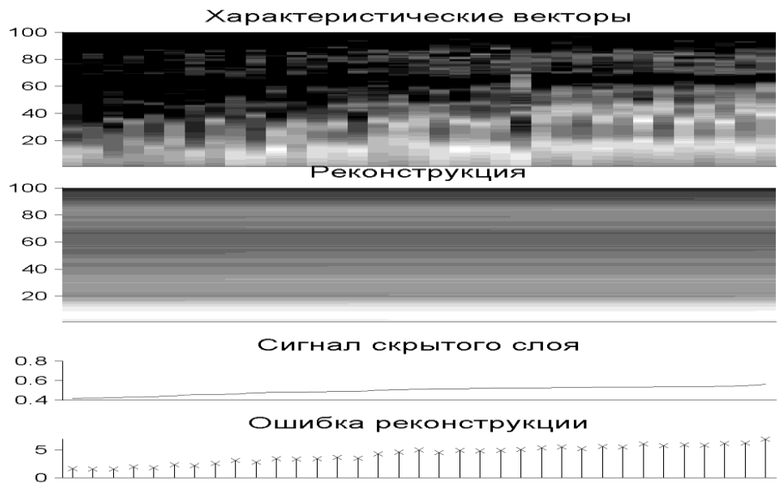
$$RL(x) = \max(0, x)$$

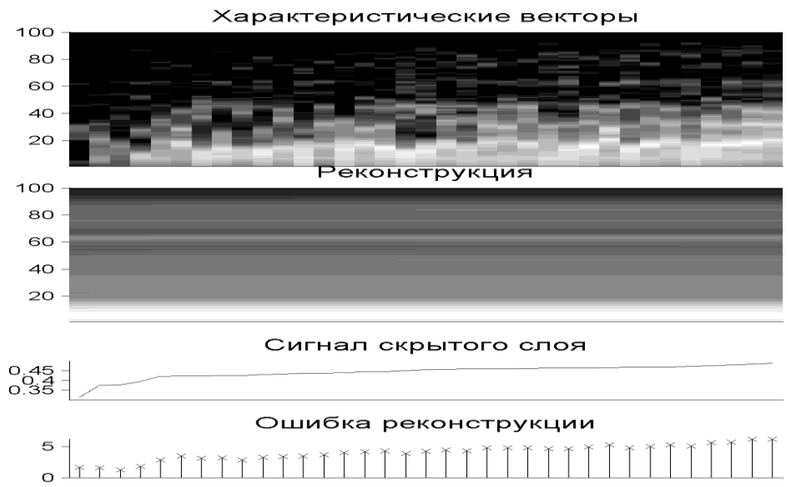
10. Процесс обучения

Женский голос

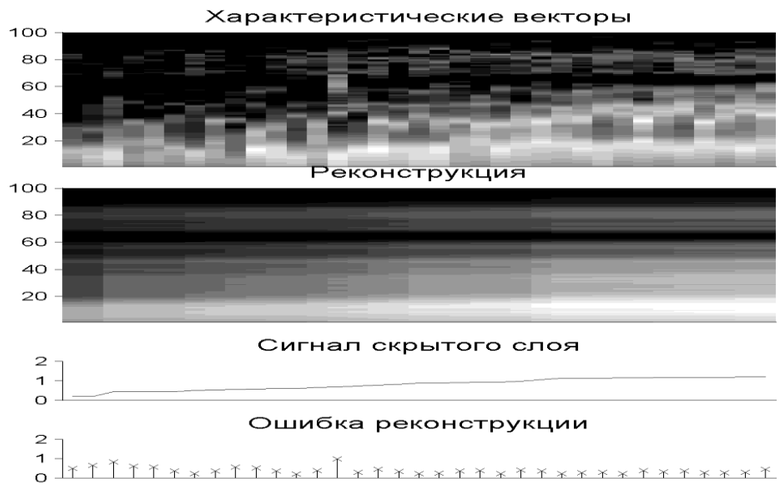
Мужской голос

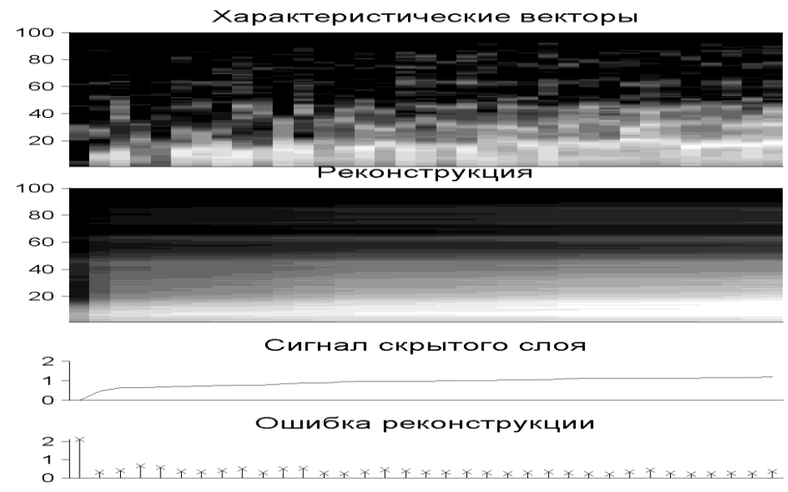
50 итераций





100 итераций



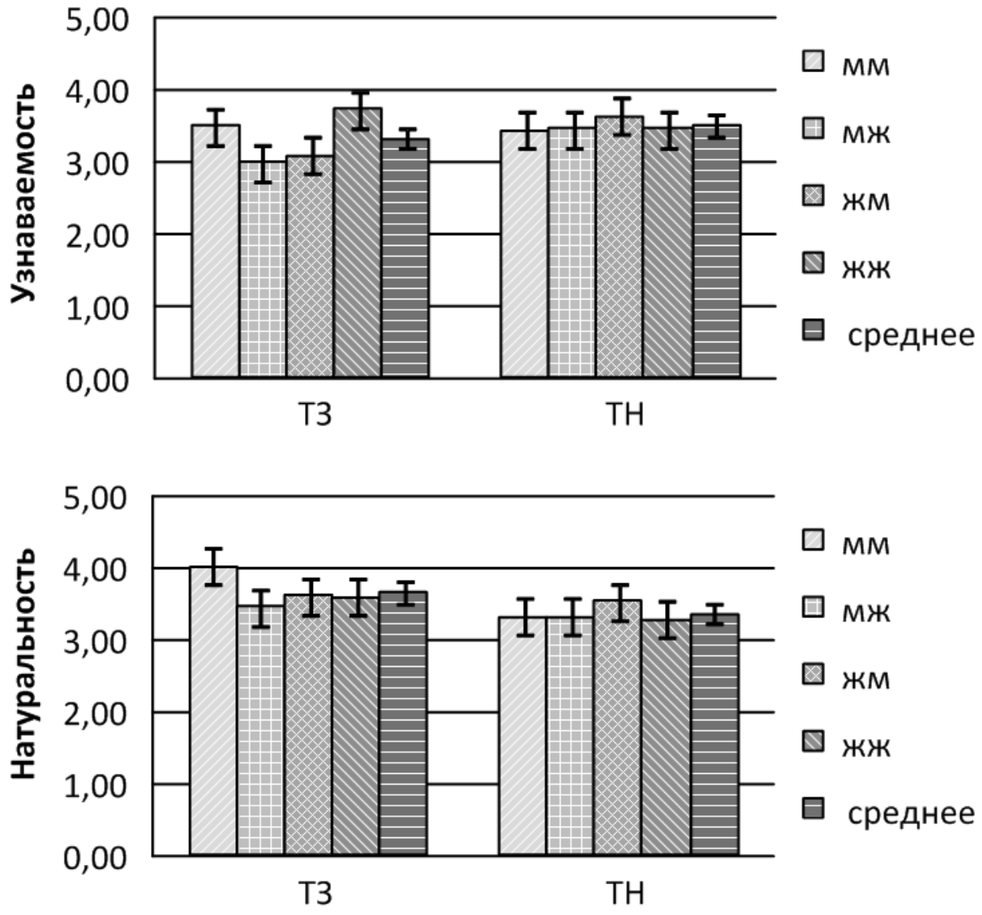


11. Конверсия голоса

- Обученные модели голосов использовались для поиска соответствия между характеристическими векторами исходного и целевого дикторов.
- Конверсия речевого сигнала выполнялась с использованием ручной фонетической разметки, в которой выделялся характерный «центральный» характеристический вектор каждой фонемы.
- На основании разметки автоматически определялись «переходные» характеристическим векторы, относящиеся к границам между фонемами.
- Из обучающей выборки целевого диктора извлекался характеристический вектор, наиболее близкий к полученному в пространстве кодов пониженной размерности.

12. Сравнение результатов

Субъективная оценка узнаваемости и натуральности конвертированной речи. Средние значения оценок экспертов.



Сравнение двух методов: 1) на основе автоматического кодера ‘ТН’;
2) на основе нейронной сети с кусочно-линейной функцией активации ‘ТЗ’.

Спасибо за внимание !!!