Ин. яз. им. М. Тореза

Новые информационные технологии

Харламов
Александр Александрович

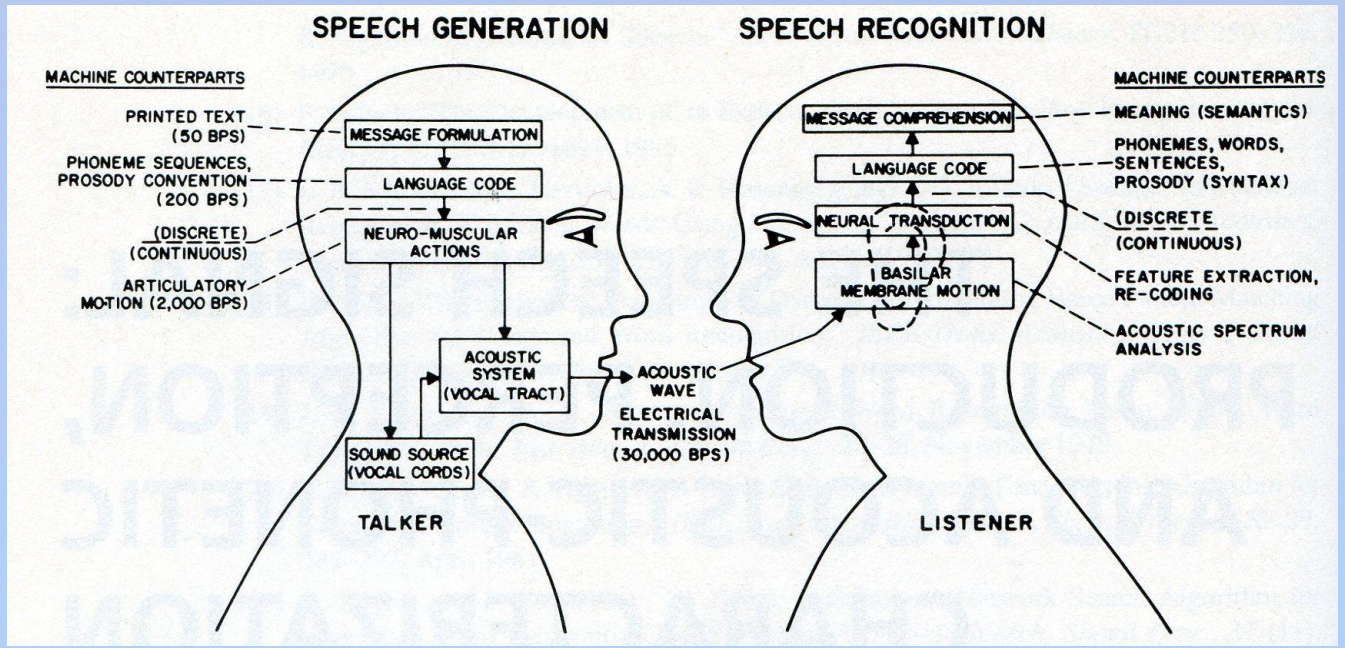
© 2012 МГЛУ

Лекция № 2. Основные классы естественно-языковых систем. Системы распознавания речи

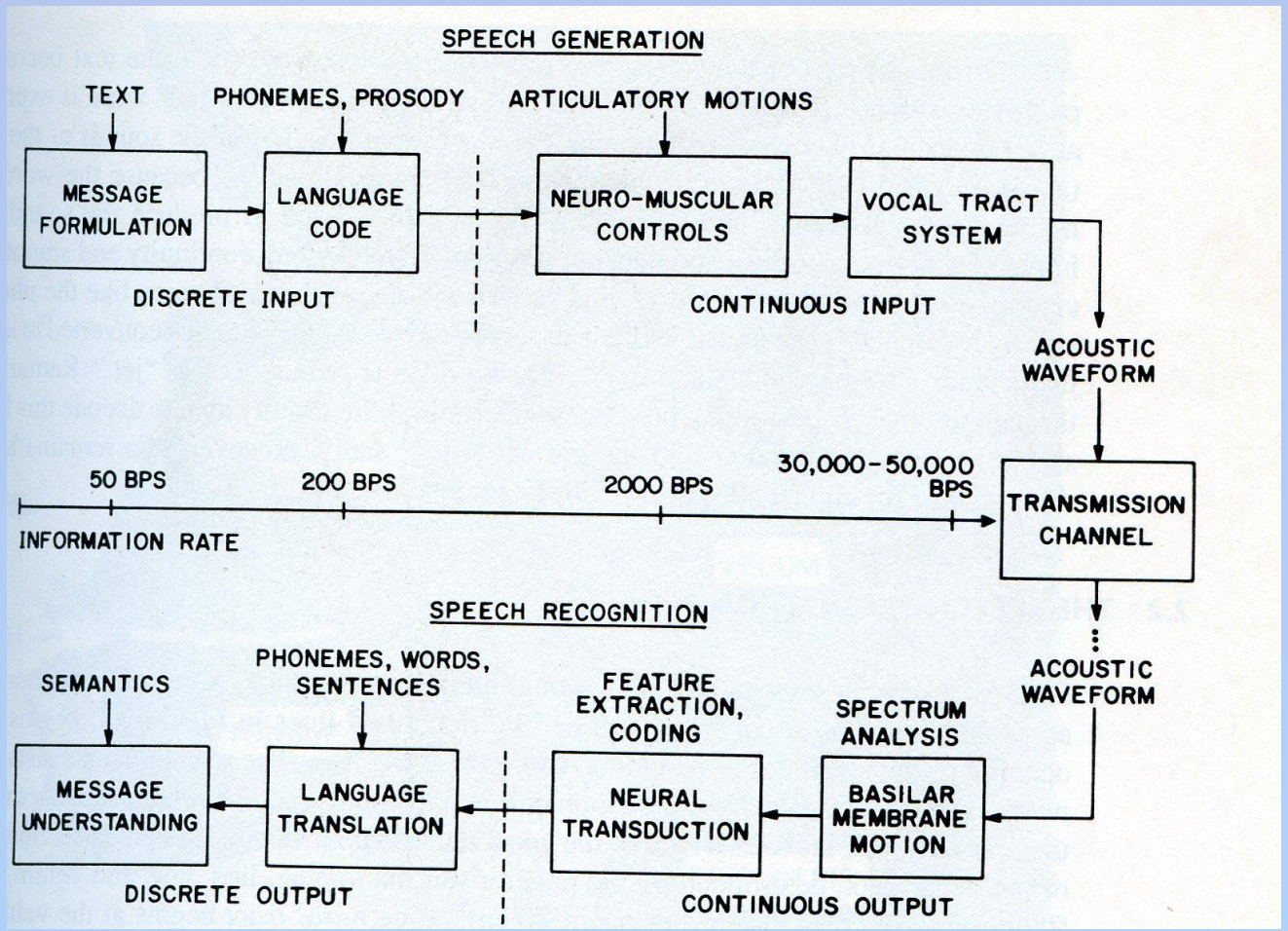
Основные классы естественно-языковых систем

Средства распознавания речи распознают голосовую (речевую) информацию и преобразуют ее в последовательность символов

Структура коммуникационной системы для организации речевого поведения

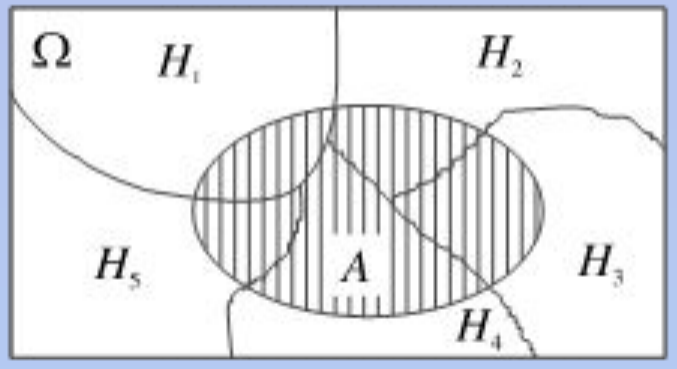


Структура коммуникационной системы для организации речевого поведения



Информационно-кодовая модель коммуникации Шеннона и Уивера

Правило Байеса



Адресант передает, а адресат принимает одно из группы событий $H_1, H_2, ..., H_n$ (классов, к которым относятся передаваемые и получаемые сообщения). Группа событий обладает следующими свойствами (она - полная):

1) все события попарно несовместны: $H_i \cap H_j = \varnothing; i, j = 1..n, i \neq j$;

2) их объединение образует пространство элементарных исходов Ω: $\Omega = H_1 \cup H_2 \cup ... H_n$

Правило Байеса

$$P(H_k / A) = \frac{P(A / H_k)P(H_k)}{\sum P(A / H_i)}$$

Пусть $H_1, H_2, ..., H_n$ - полная группа событий и $A \subset \Omega$ – некоторое событие. Тогда по формуле Байеса исчисляется вероятность реализации гипотезы H_k при условии, что событие А произошло.

Здесь А – конкретное наблюдение (измерение).

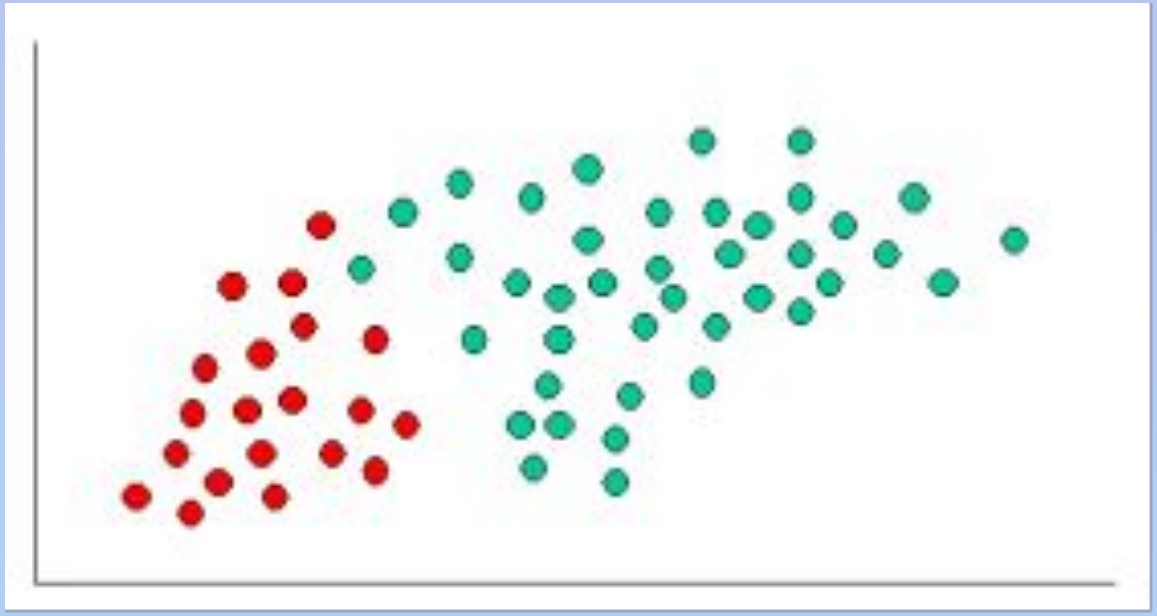
$P(H_k)$ - априорная вероятность гипотезы H_k

$P(H_k \mid A)$ - апостериорная вероятность

известны функции распределения вектора признаков для каждого класса $P(A \mid H_k)$

Правило Байеса

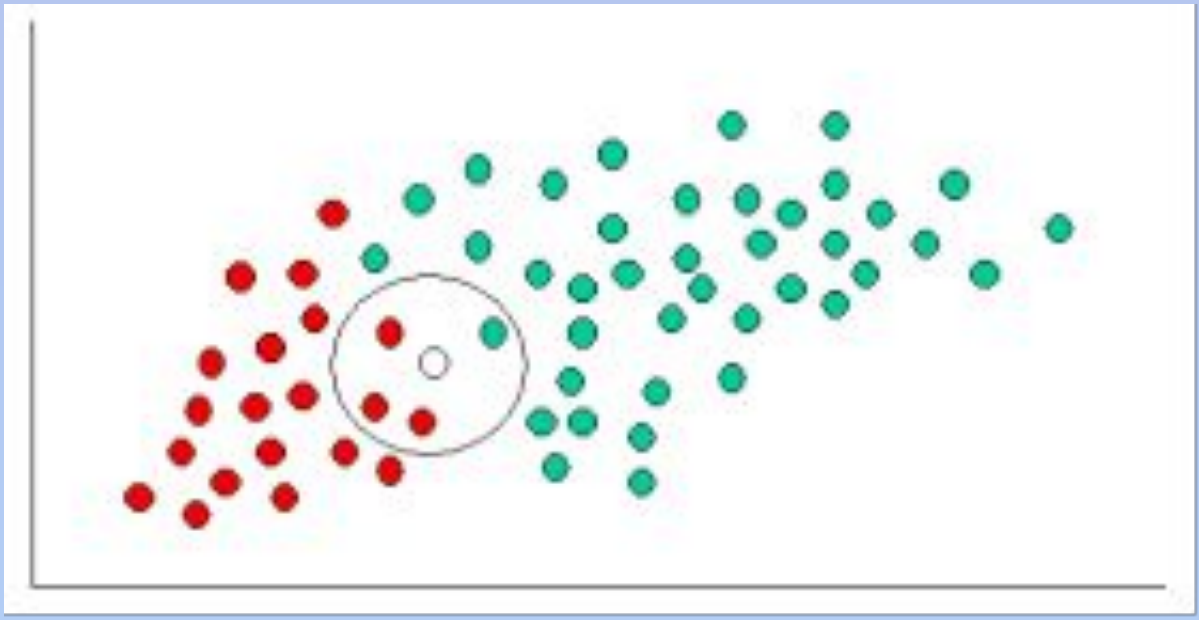
В случае двух классов H_1 и H_2, если $P(H_1 \mid A) > P(H_2 \mid A)$ то A классифицируется в H_1, иначе в H_2.



$$\text{Prior probability for GREEN} \propto \frac{\text{Number of GREEN objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} \propto \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Правило Байеса



$$\text{Likelihood of } X \text{ given } GREEN \propto \frac{\text{Number of } GREEN \text{ in the vicinity of } X}{\text{Total number of } GREEN \text{ cases}}$$

$$\text{Likelihood of } X \text{ given } RED \propto \frac{\text{Number of } RED \text{ in the vicinity of } X}{\text{Total number of } RED \text{ cases}}$$

Правило Байеса

$$P(H_k / A) = \frac{P(A / H_k)P(H_k)}{\sum P(A / H_i)}$$

Posterior probability of X being GREEN $\propto$

Prior probability of GREEN $\times$ *Likelihood of X given GREEN*

$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$

Posterior probability of X being RED $\propto$

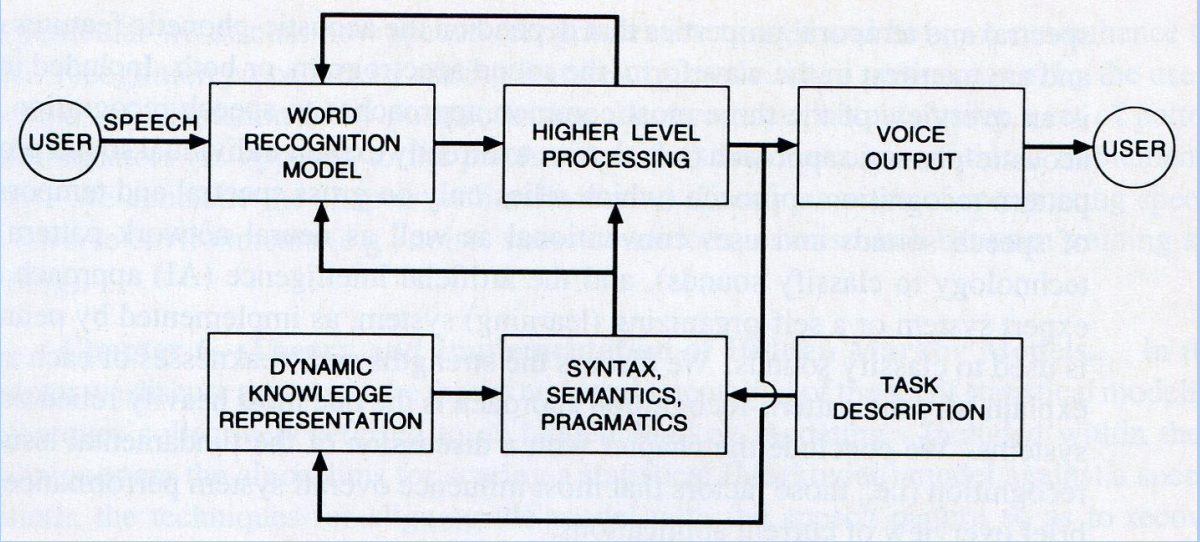
Prior probability of RED $\times$ *Likelihood of X given RED*

$$= \frac{2}{6} \times \frac{3}{40} = \frac{1}{40}$$

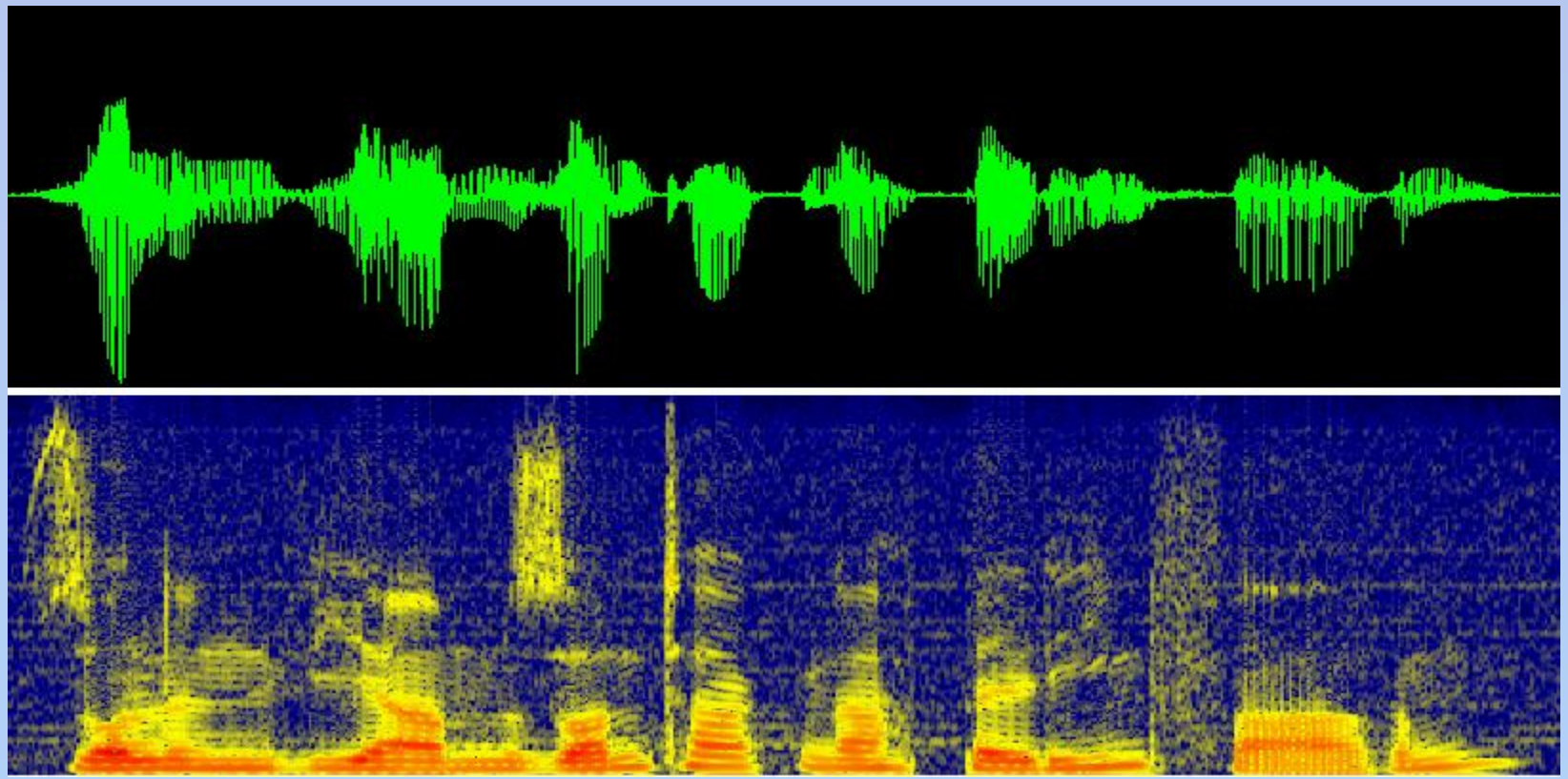
Информационно-кодовая модель коммуникации Шеннона и Уивера, модифицированная для коммуникационного акта Якобсоном

Автоматическое распознавание речи

Общая блок-схема ориентированной на задачу системы распознавания-синтеза речи



Речевая волна во временной и частотной областях



Процесс восприятия речи человеком

Структура речевой информации

- уровень семантических представлений (словарь попарной сочетаемости слов)
- синтаксический уровень (словарь синтаксем)
- лексикон (словарь корневых основ)
- морфологический уровень (словарь окончаний)
- акустико-фонетический уровень

Системы распознавания речи

1. Системы покомандного распознавания
2. Системы распознавания ключевых слов в потоке слитной речи
3. Системы распознавания связной речи
4. Системы распознавания слитной речи

Системы покомандного распознавания

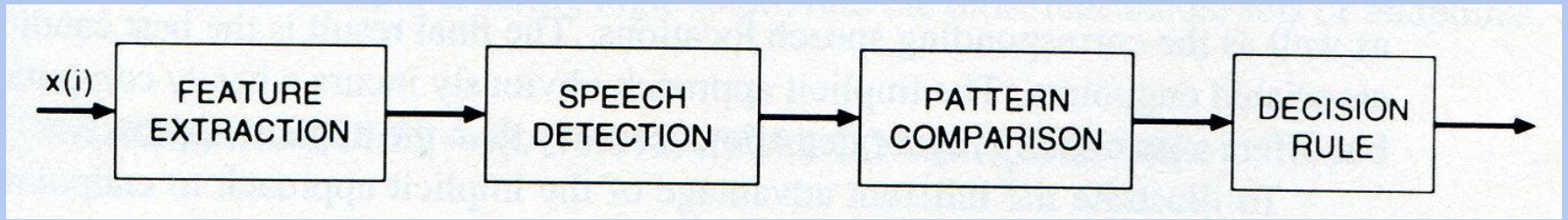
Средства голосового ввода, управления и сбора данных предназначены для ввода голосовых команд, управляющих работой некоторой системы (например бытовой техникой)

Системы покомандного распознавания

- лексикон (словарь корневых основ)
- акустико-фонетический уровень

Покомандное распознавание речи

Базовый алгоритм покомандного распознавания

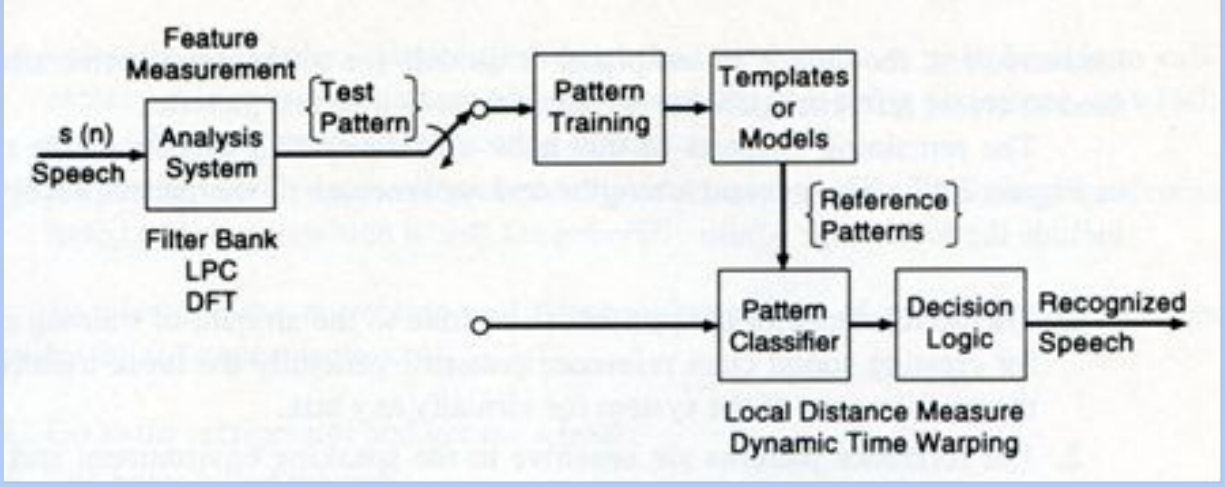


1) Формирование вектора признаков
2) Сегментация
3) Сравнение входного и эталонных образов
4) Принятие решения

Покомандное распознавание речи

Имеется три основных подхода к покомандному распознаванию речи
1. Подход, основанный на распознавании образов
2. Акустико-фонетический (структурный) подход
3. Подход, основанный на использовании искусственных нейронных сетей

Подход, основанный на распознавании образов



Блок-схема распознавателя речи на основе подхода, основанного на распознавании образов

Подход, основанный на распознавании образов

Алгоритм включает четыре основных шага:

1) Измерение первичных признаков речевого сигнала. Вычисляются спектральные признаки, либо с помощью гребенки фильтров, либо с помощью линейного предсказывающего кодирования, либо с помощью Дискретного преобразования Фурье

2) Формирование эталонов (обучение). Эталоны формируются с помощью некоторых усредняющих процедур. Это может быть модель, характеризующая статистику признаков эталона.

Подход, основанный на распознавании образов

Алгоритм включает четыре основных шага:

3) Классификация, во время которой входной образ сравнивается с эталонами. Сравнение образов, которые есть последовательность векторов признаков, осуществляется с использованием как локального расстояния между двумя синхронными векторами, так и глобальной выравнивающей во времени процедуры (чаще всего, это процедура динамического программирования), которая компенсирует различные скорости произнесения входного и эталонного образов

4) Принятие решения

Подход, основанный на распознавании образов

Достоинства подхода:

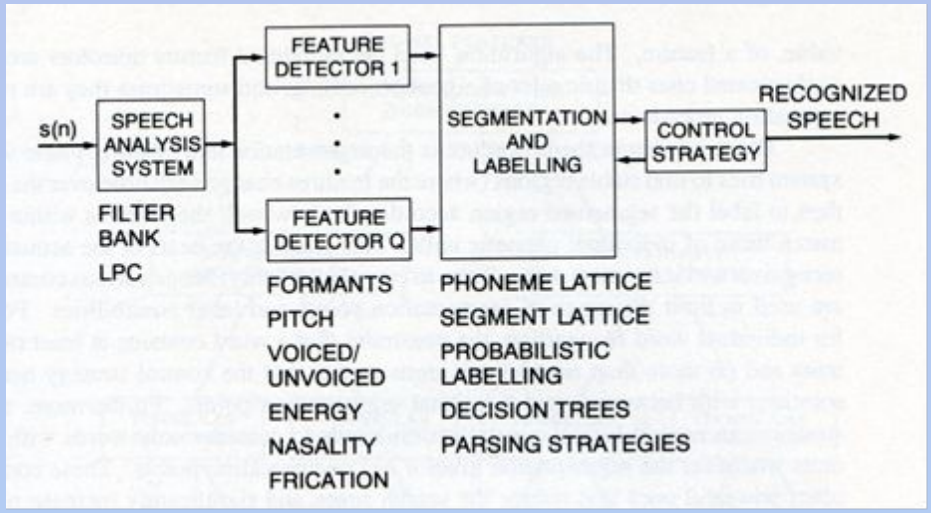
1. Простота. Он легко интерпретируется
2. Имеется хорошо развитый математический аппарат для всех процедур подхода
3. Устойчивость и инвариантность к различным словарям, пользователям, выбору признаков, использования алгоритмов сравнения образов и принятия решения, а также групп дикторов, используемого оборудования, канала
4. Не зависит от выбора речевой единицы: от фонемы до фразы. Дает хорошие результаты в широком круге задач

Подход, основанный на распознавании образов

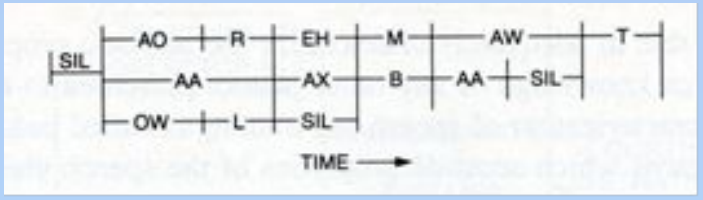
Недостатки:

1. Чувствительность к объему обучающей выборки
2. Подверженность качества распознавания влиянию шума
3. Неиспользование лингвистической информации
4. Большая вычислительная емкость

Акустико-фонетический подход



Фонетические гипотезы, полученные при распознавании строки слов «all about»



Акустико-фонетический подход

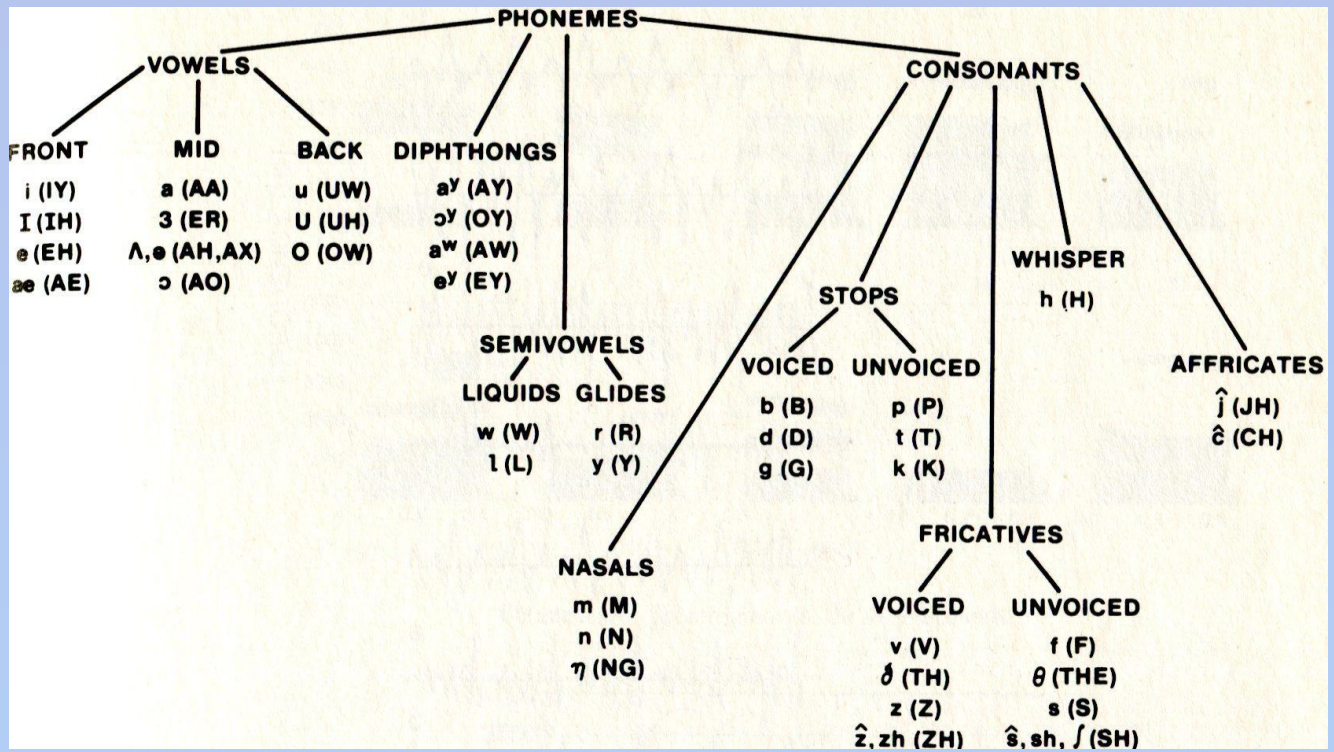
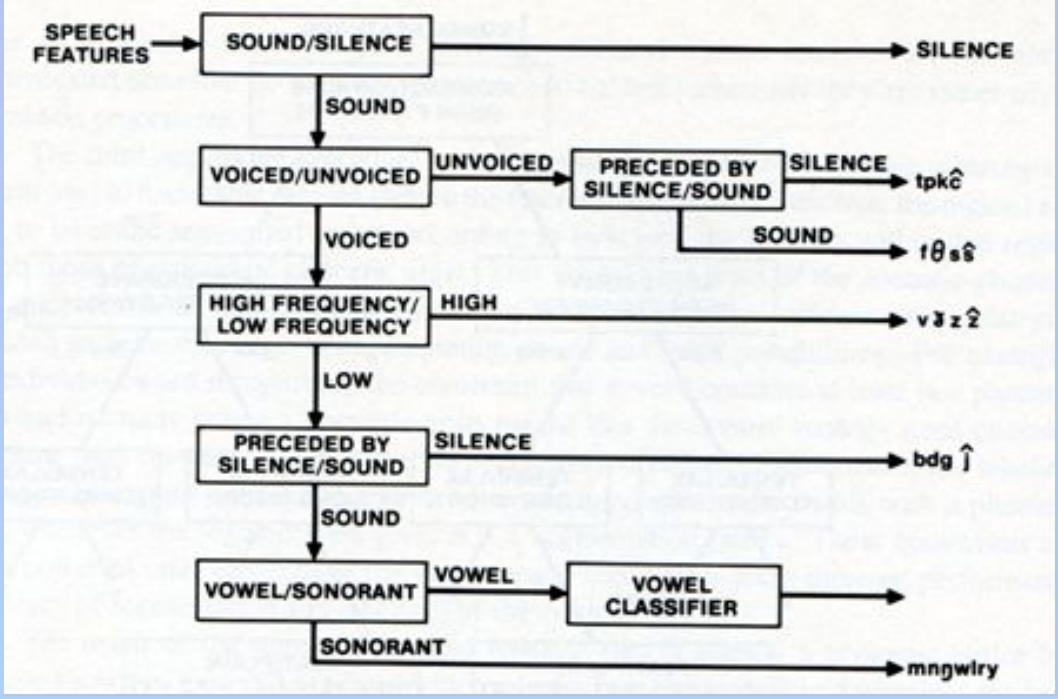


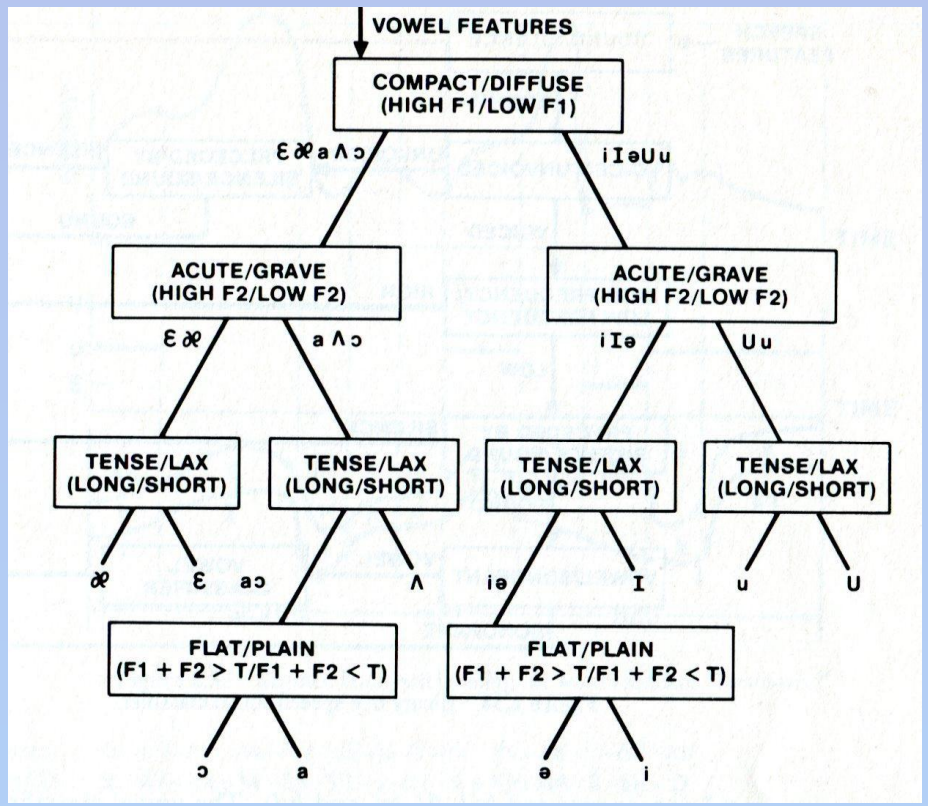
Диаграмма классификации стандартных фонем американского английского в широкие звуковые классы

Акустико-фонетический подход



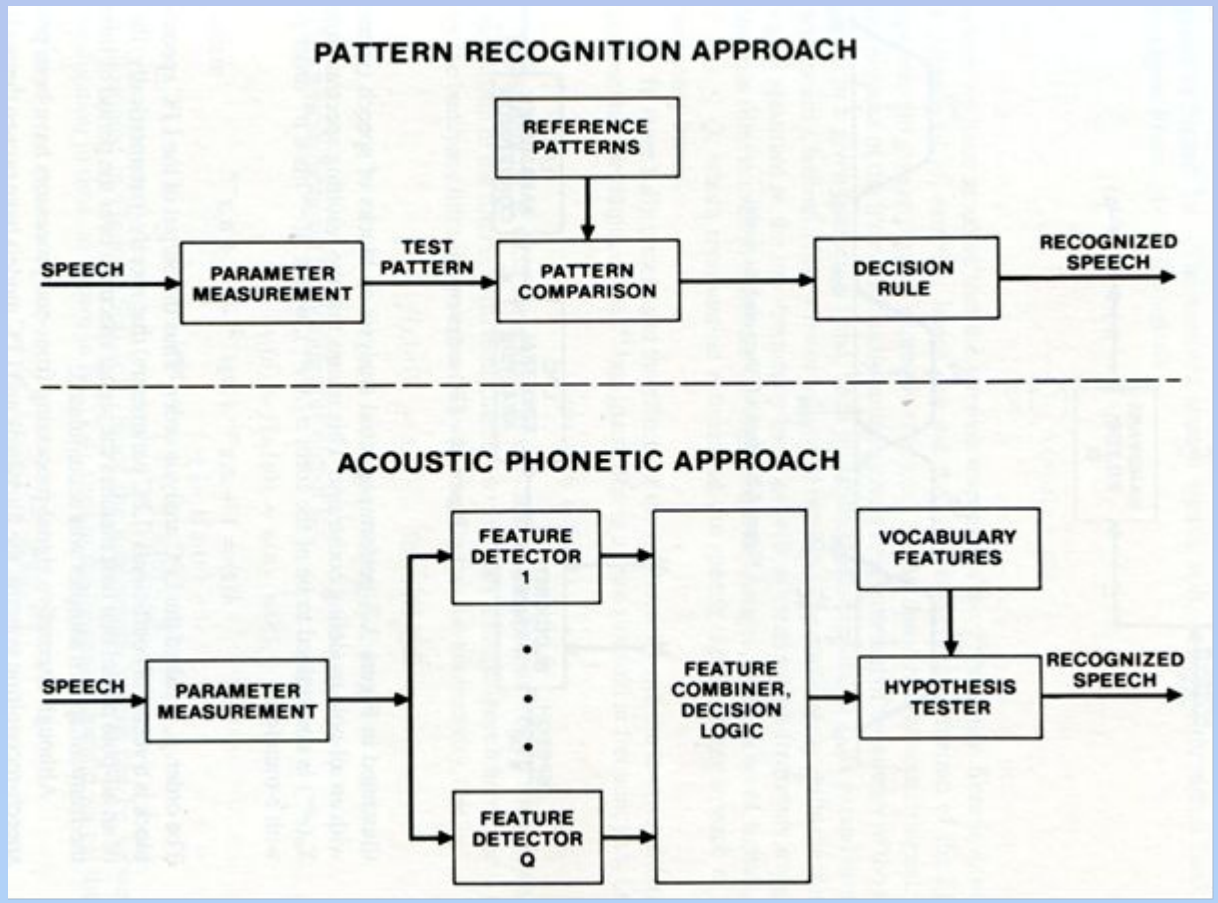
Бинарное дерево классификации речевых звуков

Акустико-фонетический подход

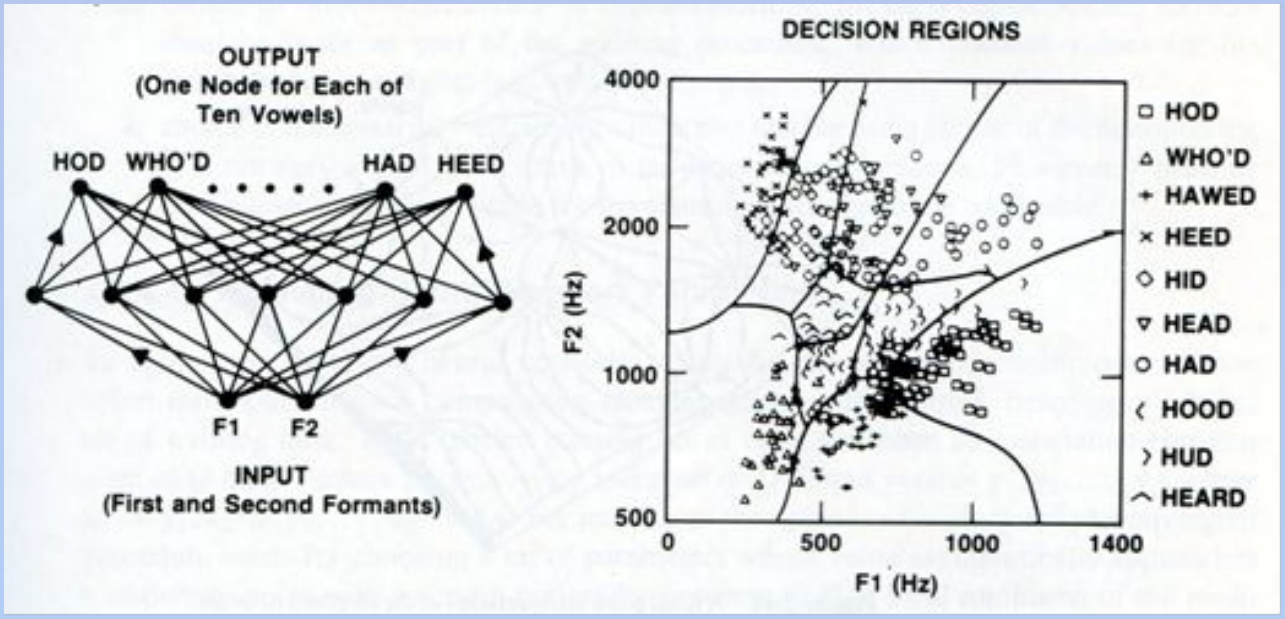


Акустико-фонетический классификатор гласных

Сравнение подходов, основанного на распознавании образов и на основе акустико-фонетического анализа



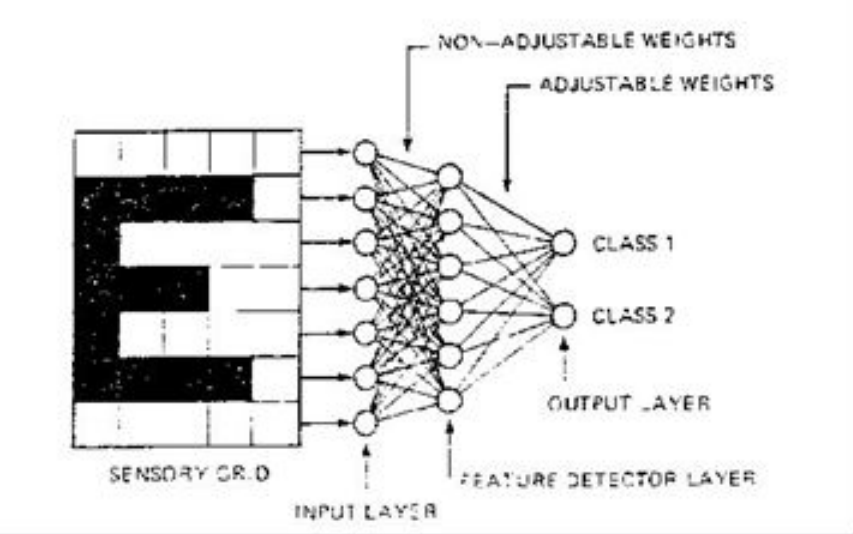
Подход, основанный на искусственных нейронных сетях



Многослойный персептрон для классификации гласных, основанной на формантных измерениях

Искусственные нейронные сети

Трехслойный персептрон



$$S_j = \sum_{i=o}^{n} a_i w_{ji}$$

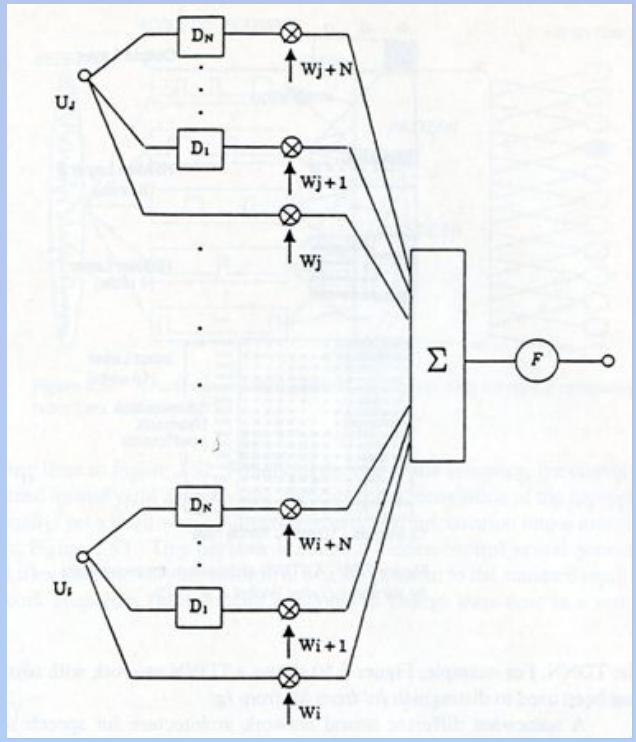
- вычисляет сумму

$$if S_j > 0 \mid x_j = 1$$

$$if S_j \le 0 \mid x_j = 0$$

- сравнивает с порогом

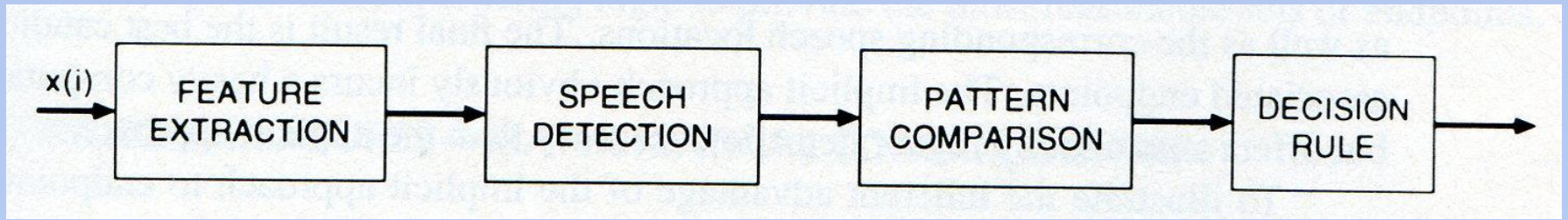
Подход, основанный на искусственных нейронных сетях



Нейронная сеть с задержками

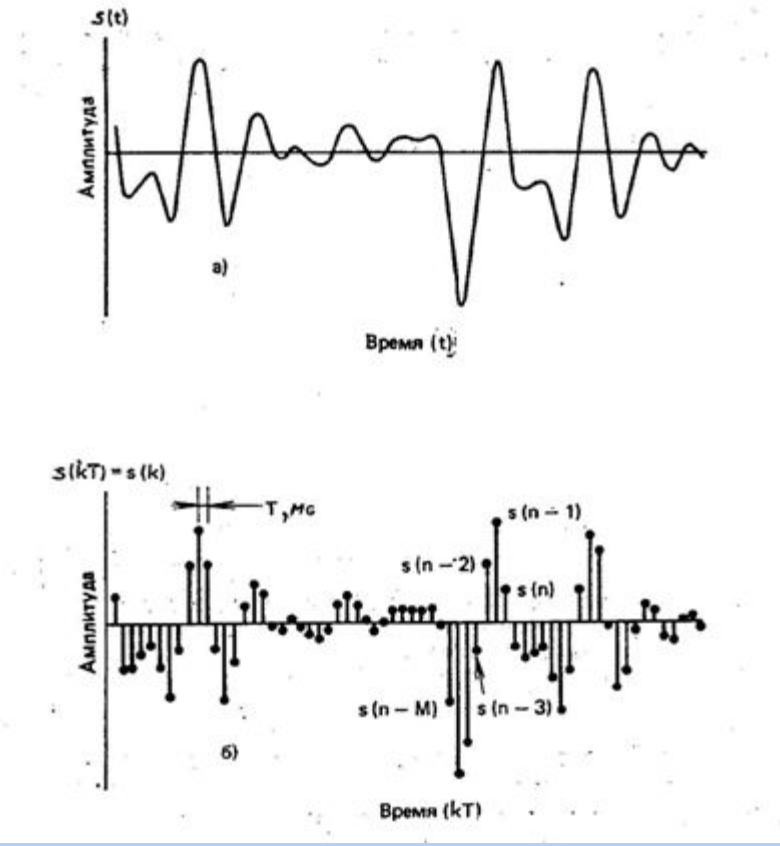
Покомандное распознавание речи

Базовый алгоритм покомандного распознавания



1) Формирование вектора признаков
2) Сегментация
3) Сравнение входного и эталонных образов
4) Принятие решения

Оцифровка речевого сигнала

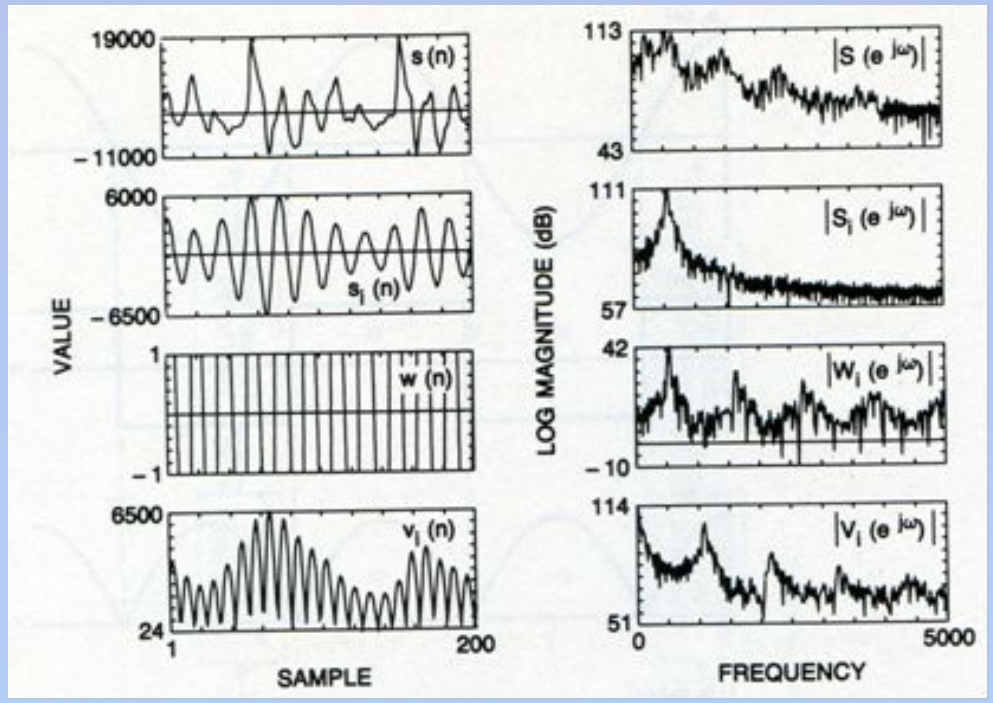


Первичная обработка

Наиболее характерные подходы:

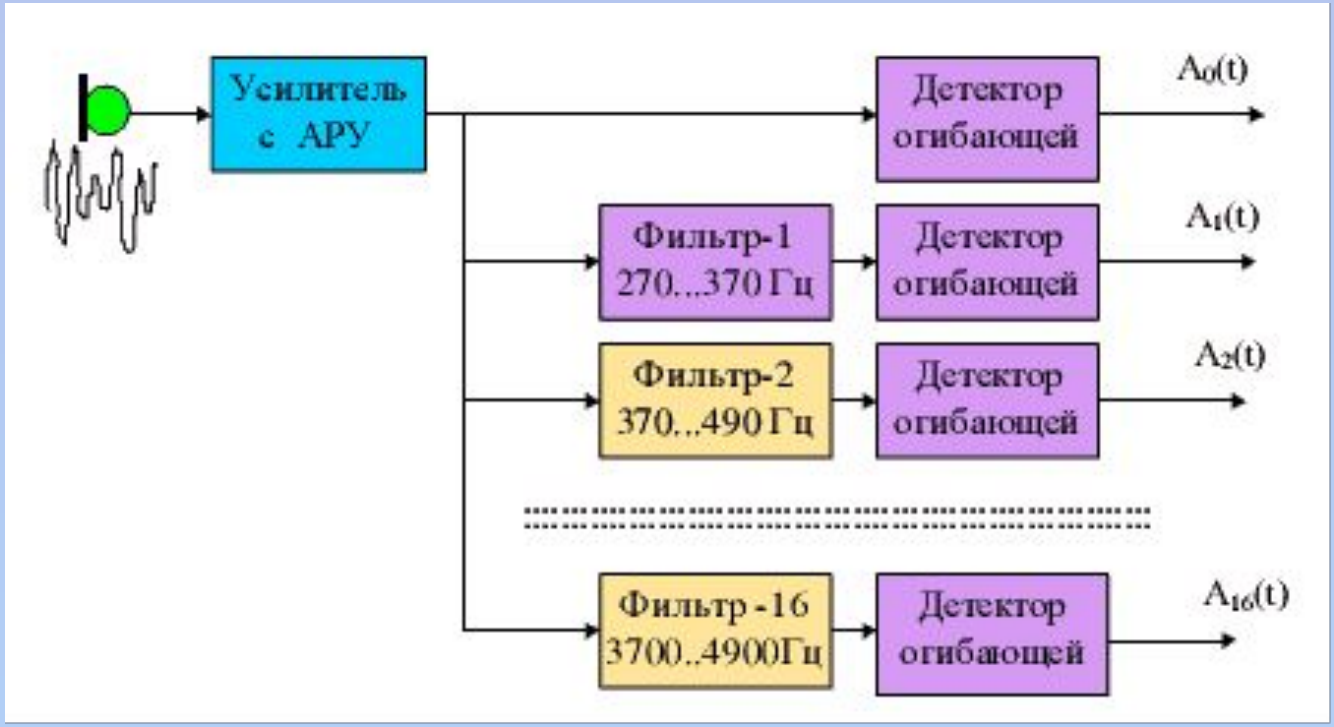
1. Спектральный анализ
2. Антропоморфная модель

Спектральный анализ



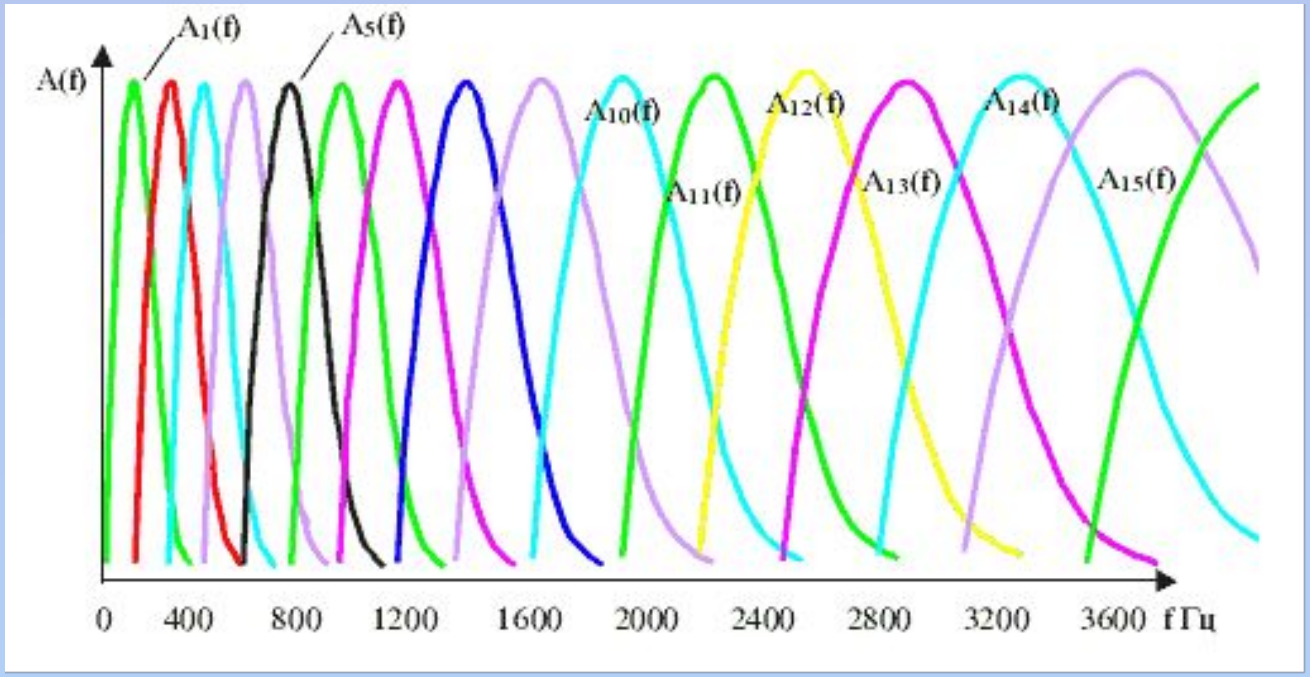
Типичный вид речевой волны и ее спектра в модели анализа на основе гребенки фильтров

Спектральный анализ



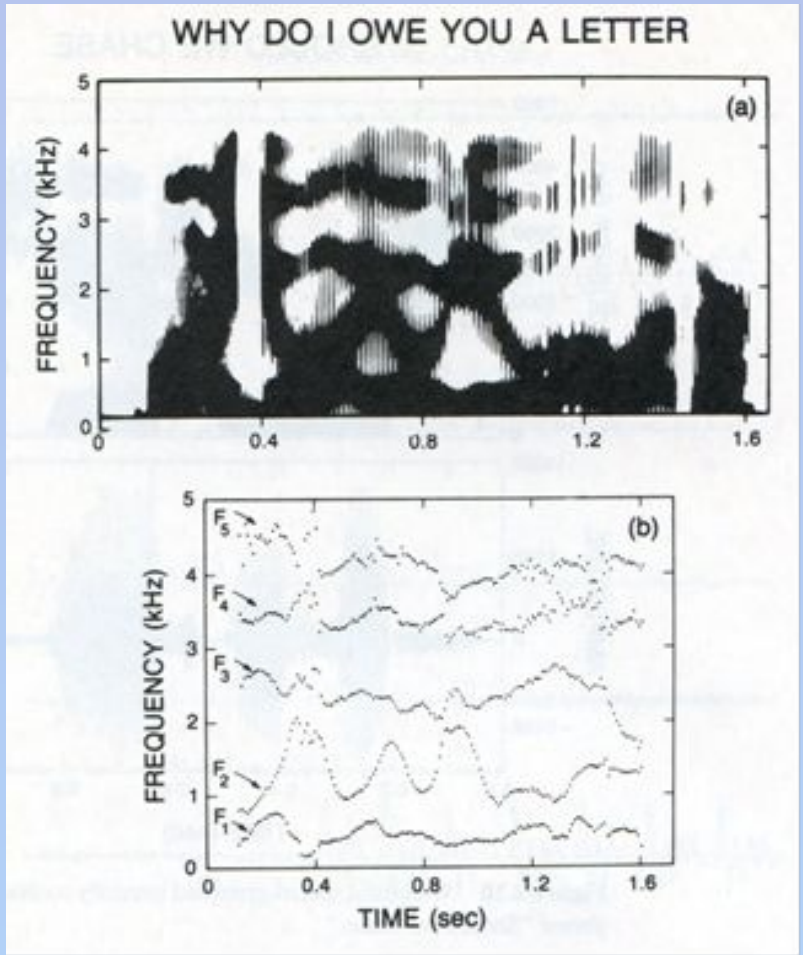
На основе гребенки фильтров

Спектральный анализ



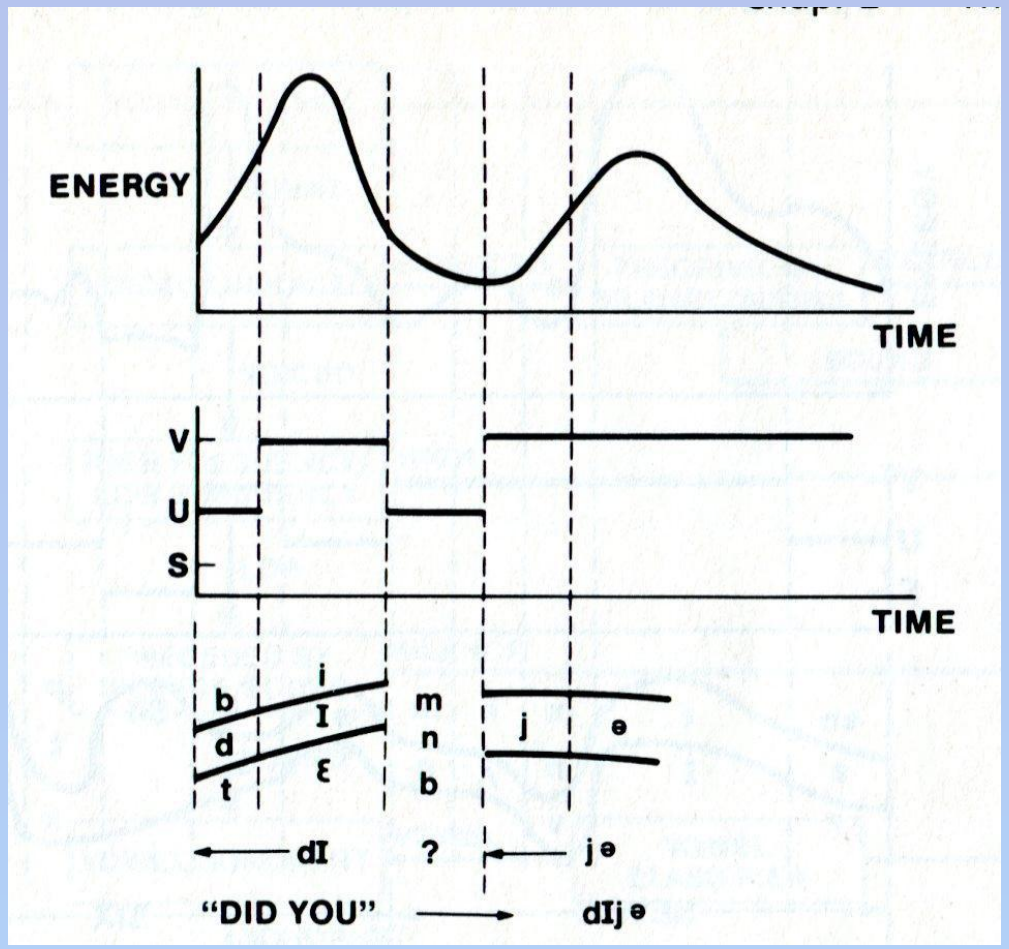
На основе гребенки фильтров

Представление речи в виде формантных траекторий



Для последовательности «Why do I owe you a letter»

Сегментация



Сегментация и разметка для последовательности «did-you»

Принятие решения

1. Динамическое программирование
2. Байесовское правило
3. Скрытые Марковские модели

Принятие решения

Ключевым вопросом в распознавании речи является вопрос сравнения входного образа с эталонными образами с целью выяснения степени их подобия

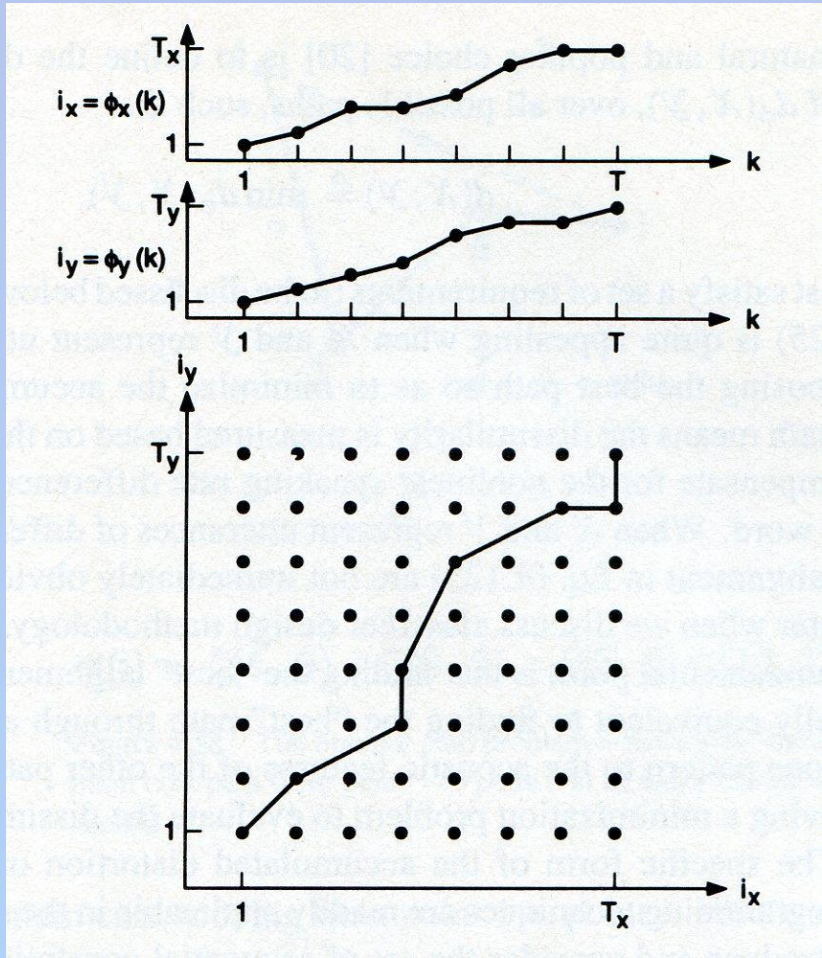
Обычно входной и эталонный образы имеют разную длину

Принятие решения

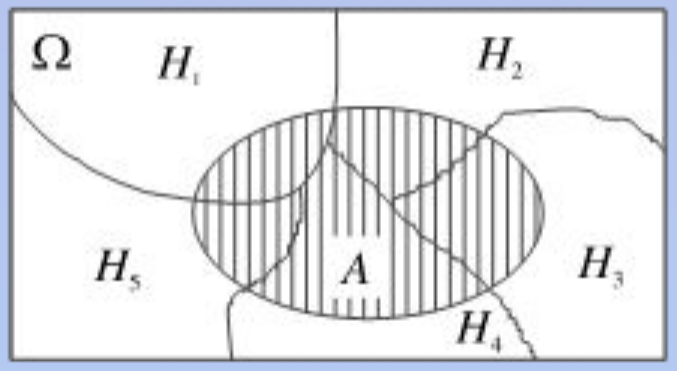
Сравниваемые образы не могут быть нормализованы по длительности, поскольку различные типы звуков имеет различные возможности по удлинению произнесения

Необходимо найти способ сравнения спектральных векторов такой, чтобы можно было вычислять глобальную меру совпадения образов

Динамическое программирование



Правило Байеса



Адресант передает, а адресат принимает одно из группы событий $H_1, H_2, ..., H_n$ (классов, к которым относятся передаваемые и получаемые сообщения). Группа событий обладает следующими свойствами (она - полная):

1) все события попарно несовместны: $H_i \cap H_j = \varnothing; i, j = 1..n, i \neq j$;

2) их объединение образует пространство элементарных исходов Ω: $\Omega = H_1 \cup H_2 \cup ... H_n$

N_k

Правило Байеса

$$P(H_k / A) = \frac{P(A / H_k)P(H_k)}{\sum P(A / H_i)}$$

Эти вероятности можно оценить методами математической статистики на множестве прецедентов.

$P(H_k) \approx N_k / N$, где N_k - число прецедентов из H_k, N - общее число прецедентов. $P(A \mid H_k)$ - гистограмма распределения вектора признаков для прецедентов из класса H_k.

N_k

Системы покомандного распознавания

Классы ЕЯ систем	Уровни представления информации (функции)						
	Акустико-фонетический	Морфологический	Лексический	Синтаксический	Модель мира	Прагматический	Сравнение (классификация)
Системы распознавания речи							
Системы покомандного распознавания	+	-	+	-	-	-	+

Системы распознавания слитной речи

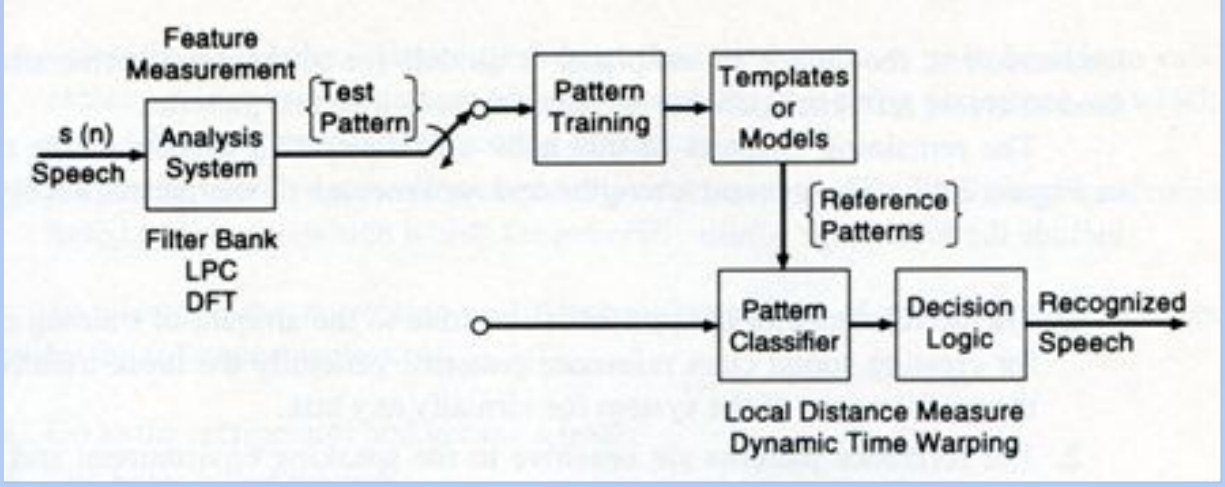
Системы типа «Речь-текст» предназначены для распознавания слитной речи (например для диктовки деловой корреспонденции)

Распознавание слитной речи

Имеется два подхода к распознаванию слитной речи:

1. Подход, основанный на распознавании образов
2. Структурный подход

Подход, основанный на распознавании образов



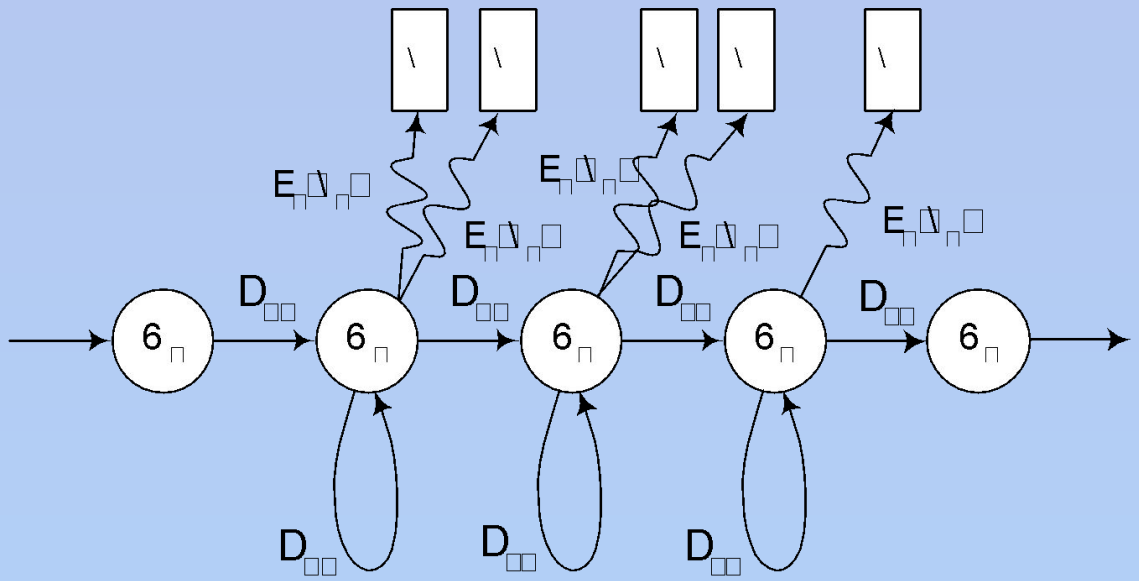
Блок-схема распознавателя речи на основе подхода, основанного на распознавании образов

Системы распознавания слитной речи

- уровень семантических представлений (словарь попарной сочетаемости слов)
- синтаксический уровень (словарь синтаксем)
- лексикон (словарь корневых основ)
- морфологический уровень (словарь окончаний)
- акустико-фонетический уровень

Подход, основанный на распознавании образов

Модель фонемы

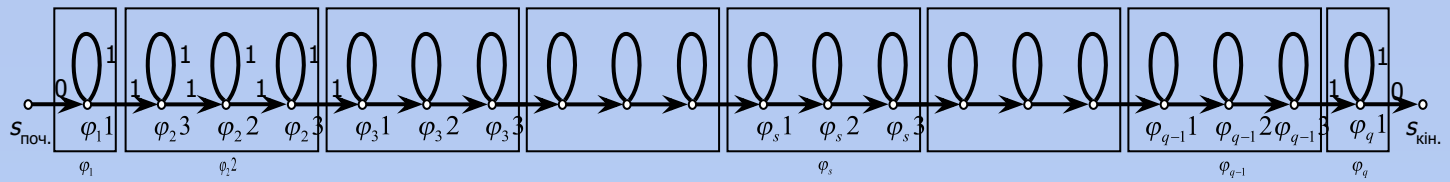


Подход, основанный на распознавании образов

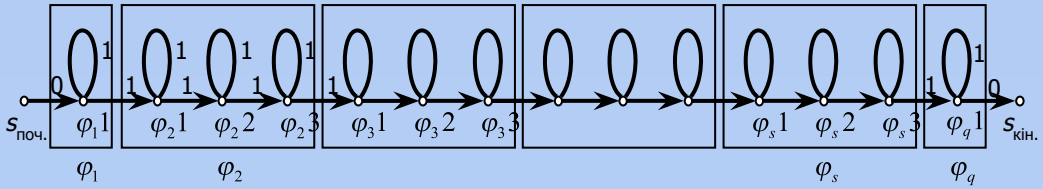
Составление эталонных сигналов слов из фонем в соответствие с моделью произношения

$$k(X_{0l}) = \underset{k}{\operatorname{argmax}} \max_{\substack{(\mu_{ks}, w_{k1s}, w_{k2s}), \\ s=1:q_k}} \sum_{s=1}^{q_k} \{(\ ) + (\ ) + (\ )\}$$

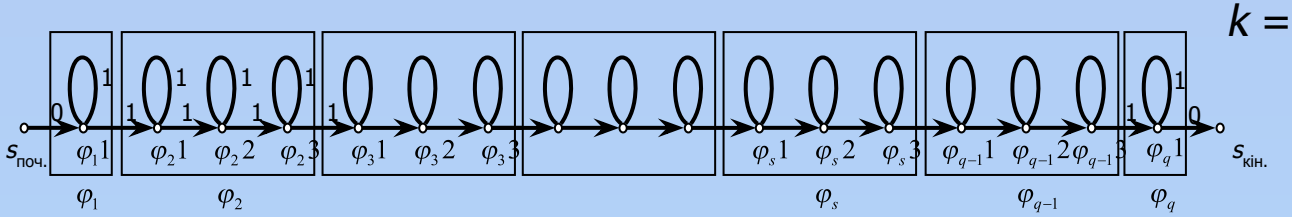
$k = 1$



$k = 2$

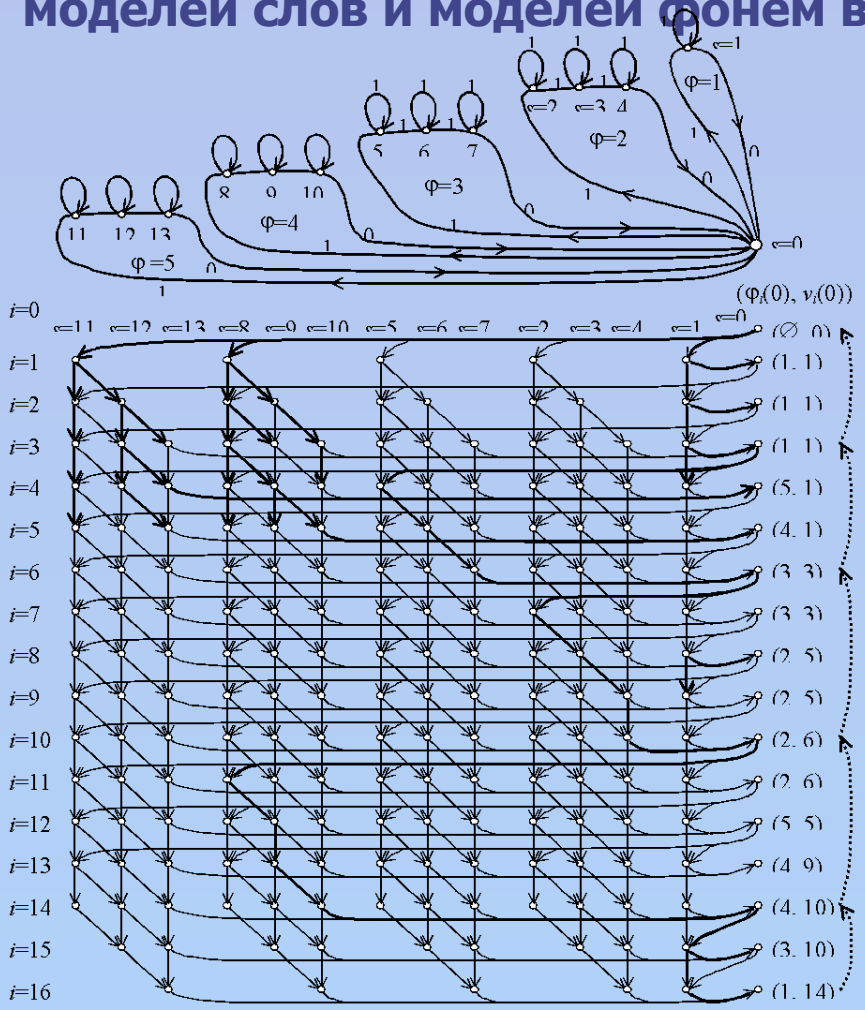


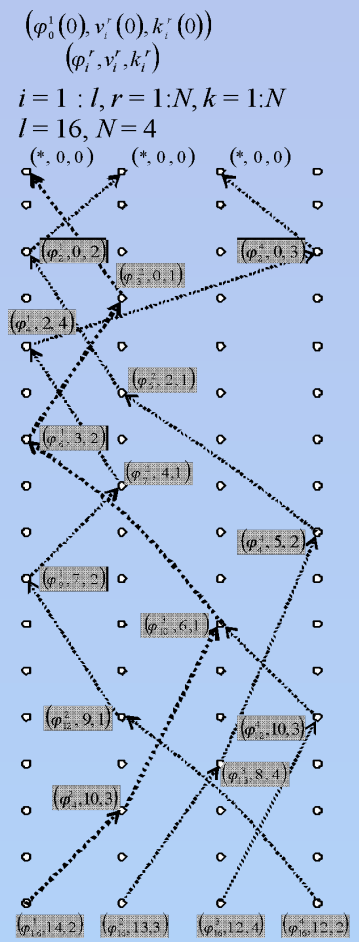
$k = 3$



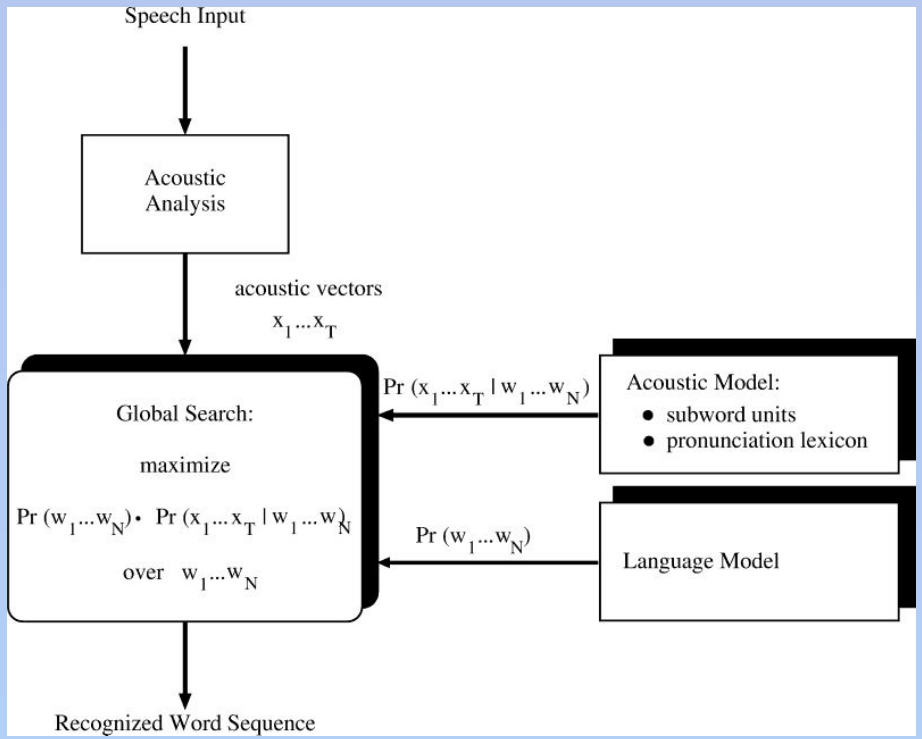
Подход, основанный на распознавании образов

Модель предложения с произвольным порядком следования моделей слов и моделей фонем в предложении



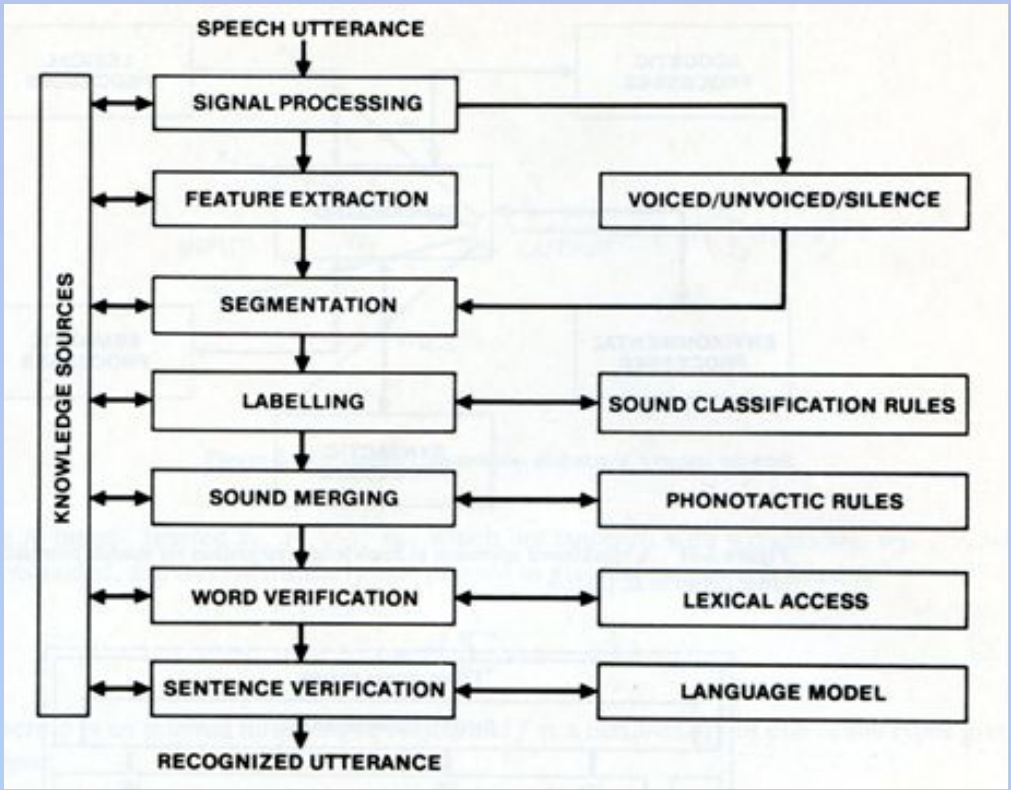


Подход, основанный на распознавании образов



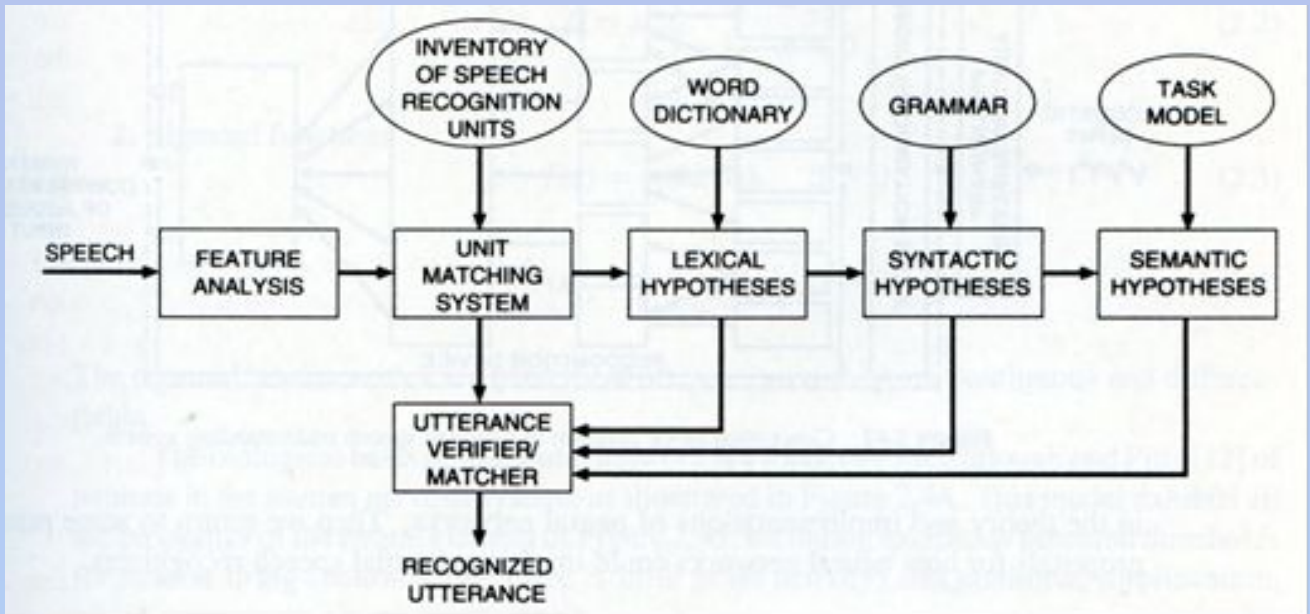
Блок-схема распознавателя речи на основе подхода, основанного на распознавании образов

Структурный подход



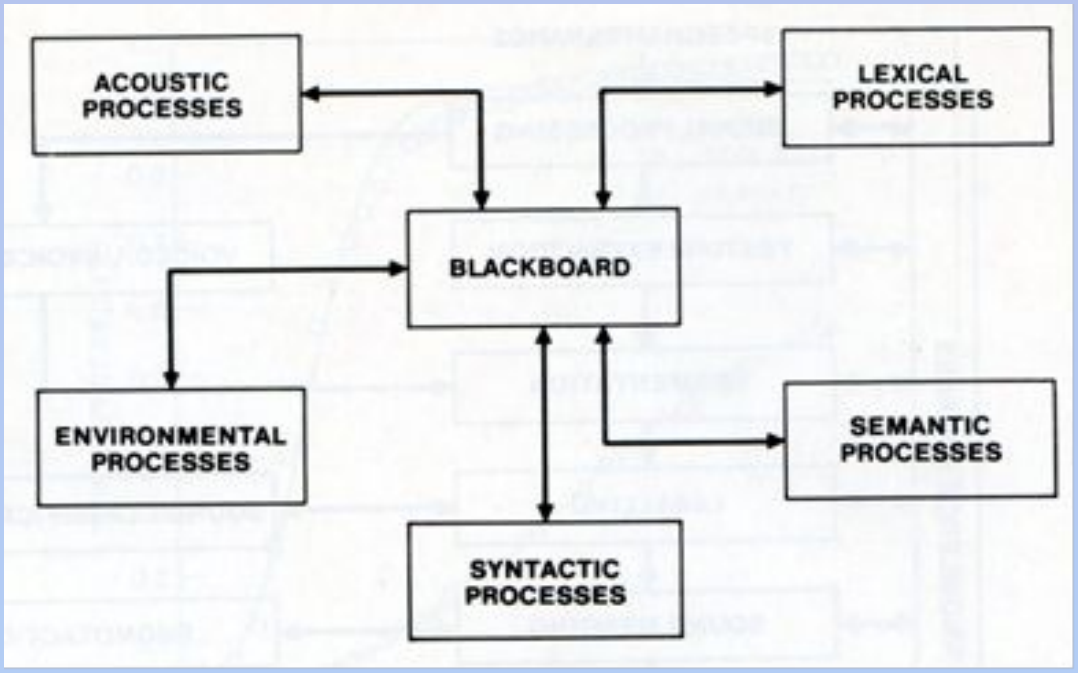
«Снизу-вверх» подход интеграции знаний для Распознавания слитной речи

Структурный подход



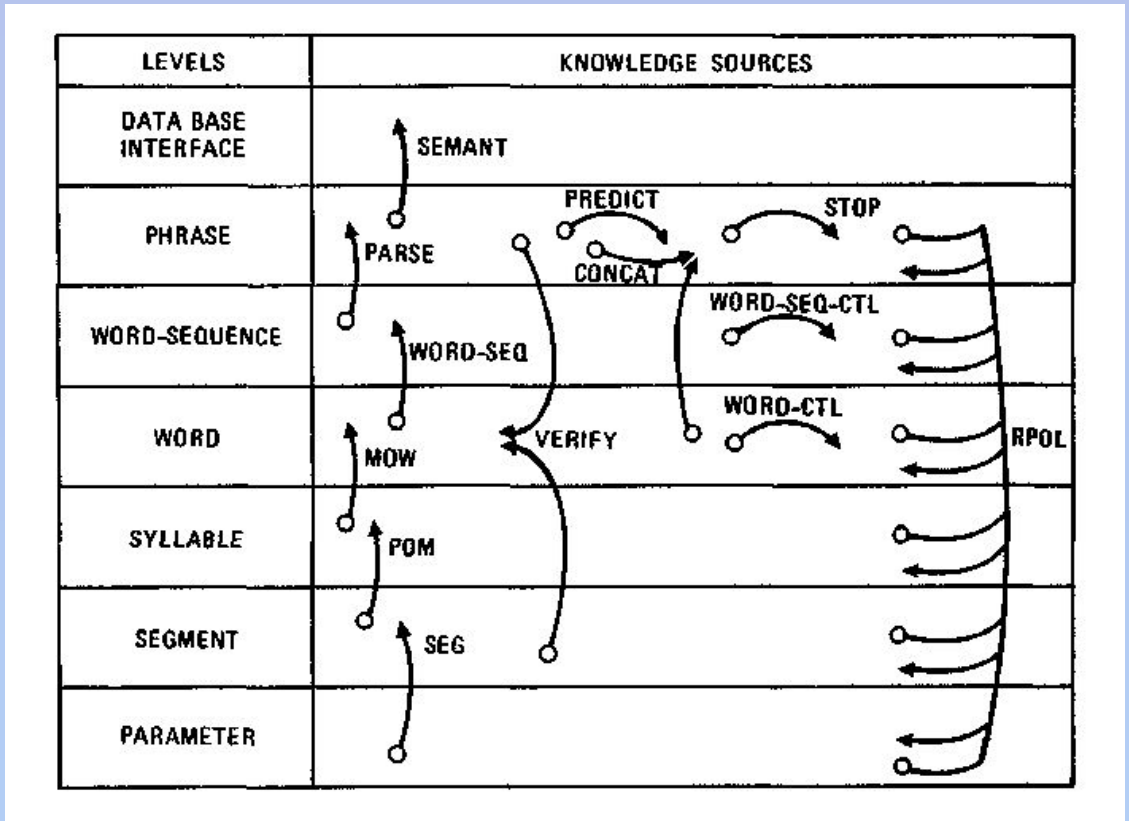
«Сверху-вниз» подход к интеграции знаний для распознавания слитной речи

Структурный подход



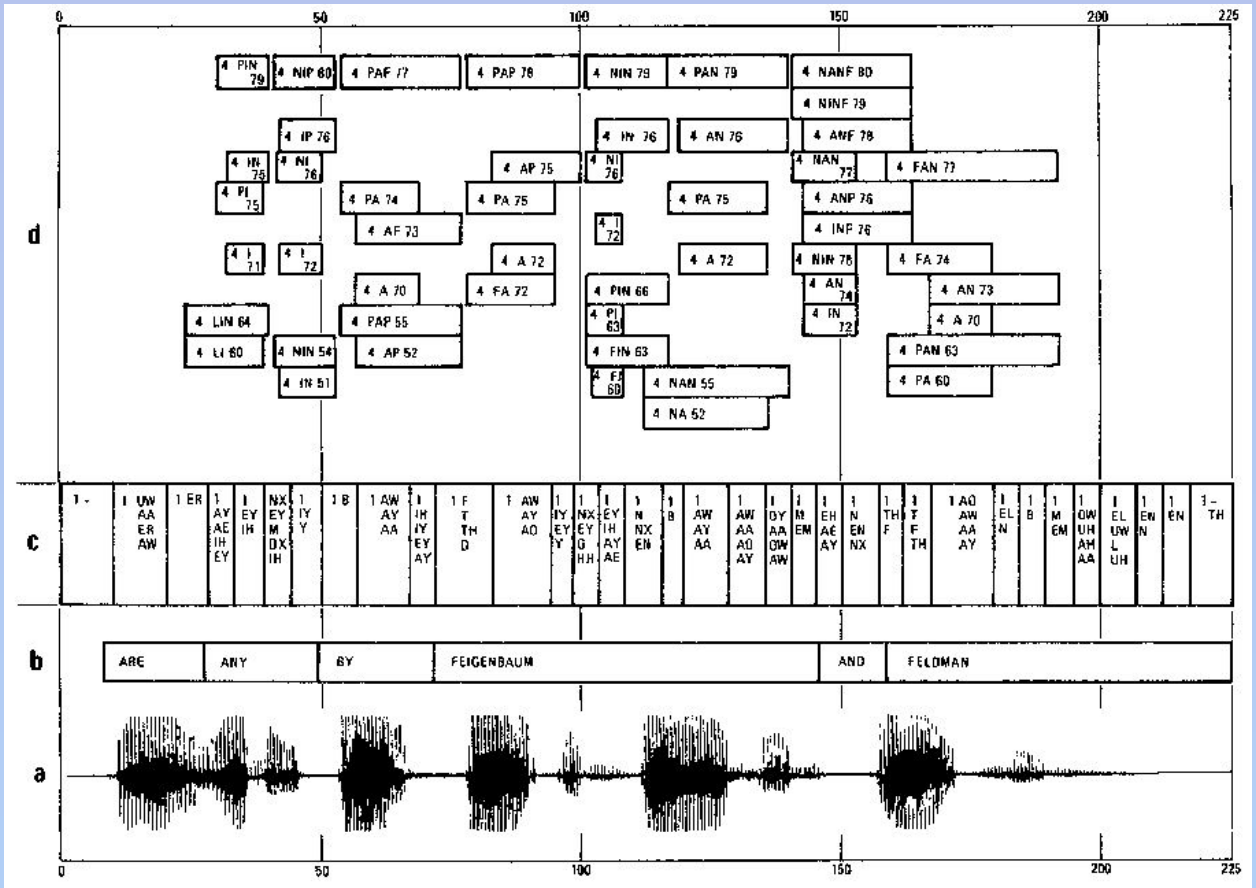
Подход, основанный на концепции классной доски для интеграции знаний

Структурный подход



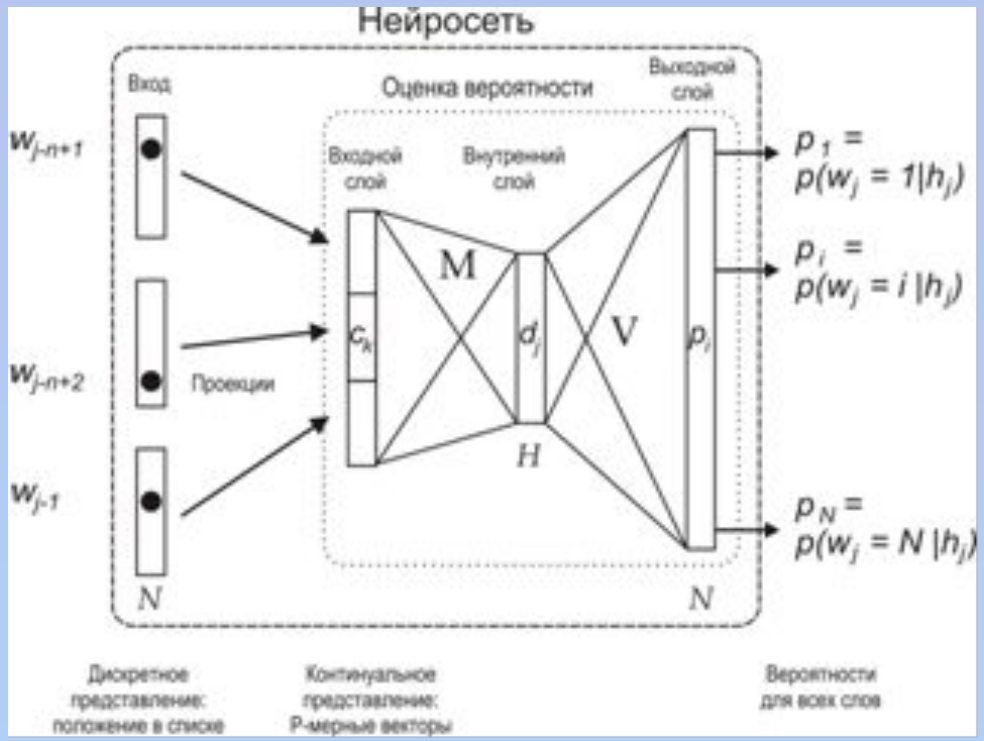
Взаимодействие между источниками знаний разных уровней

Структурный подход



Частичные гипотезу на разных уровнях

Структурный подход
(синтаксический уровень)



Модель языка на основе многослойного персептрона

Системы распознавания слитной речи

Классы ЕЯ систем	Уровни представления информации (функции)						
	Акустико-фонети-ческий	Морфоло-гический	Лекси-ческий	Синтакси-ческий	Модель мира	Прагма-тический	Сравнение (классификация)
Системы распознавания речи							
Системы распозна-вания слитной речи	**+**	**+**	**+**	**+**	**+**	**-(+)**	**+**

Сравнение эффективности распознавания человеком и искусственными системами

Задача	Процент ошибок человека	Процент ошибок ИСРР
База "TI46", SNR ~60 дБ Звуки, обозначающие буквы алфавита	1.6% [1]	5% [2]
База "TI digits", SNR ~60 дБ Цифровые последовательности	0.009%	0.72%
	0.105% [3]	
База "Resource Management", SNR ~60 дБ (словарь 1000 слов, языковое моделирование)	0.1%	3.6%
База "Resource Management", SNR ~60 дБ (словарь 1000 слов, нет модели языка)	2%	17%
База "Wall Street Journal", SNR ~60 дБ (словарь 5000 слов, чтение)	0.9%	7.2%

Сравнение эффективности распознавания человеком и искусственными системами

(продолжение)

База "Switchboard", SNR ~60дБ (задача распознавания спонтанной речи)	4%	43%
База “Corpus of Spontaneous Japanese”, SNR ~ 60 дБ (задача распознавания спонтанной речи)	4%	9%
База "Switchboard", SNR ~60 дБ (задача выделения 20 ключевых слов)	12.8% [4] 7.4% [5]	31.1 %
Слитная речь, SNR ~60 дБ (словарь 20000, чтение)	2.6%	12.6 %

1 - распознавание последовательностей
2 - распознавание изолированных слов
3 - распознавание вокодерной речи (модель линейного предсказания 12-го порядка)
4 - отсутствие контекста, бессмысленный поток слов
5 - осмысленный контекст ~ 2 сек

Сравнение эффективности распознавания человеком и искусственными системами в условиях шума

Отношение Сигнал-Шум	Процент ошибок человека	Процент ошибок ИСРР
SNR ~60 дБ	~1%	~1%
SNR 18 дБ	~1%	~10%
SNR 12 дБ	~1%	~25%
SNR 6 дБ	~1%	~60%
SNR 0 дБ	~1%	~100%

Диалог человека и машины

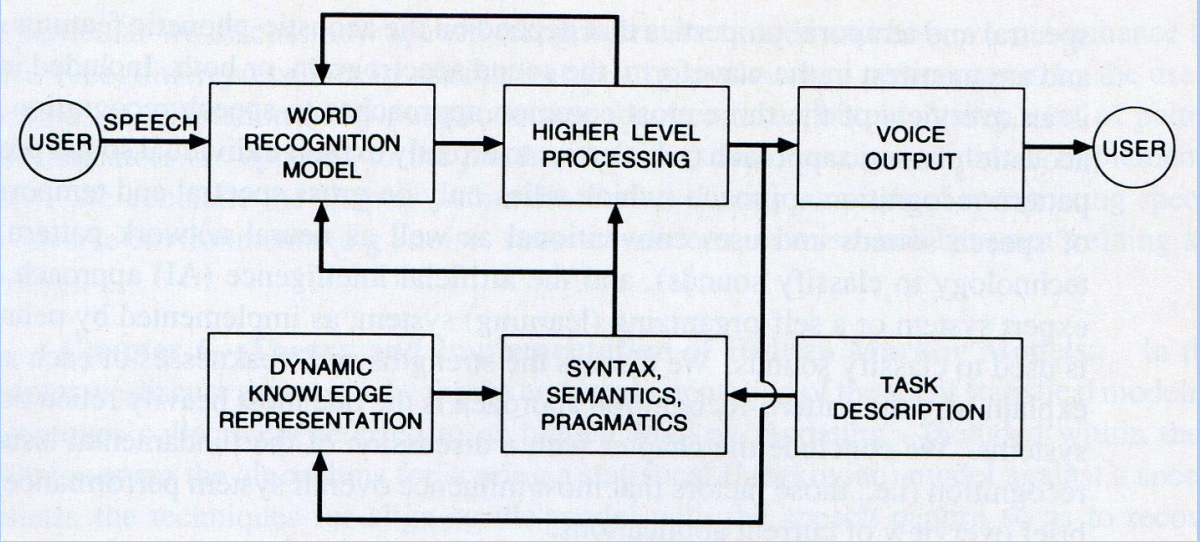
Существует два типа таких систем

Системы первого типа ведут человека, следуя четкому порядку заполнения определенных форм. Примером такой системы может стать заказ билетов. В каждом из состояний система настраивается на распознавание заранее приготовленных атрибутов, допуская варианты возврата к предыдущему шагу или выходу из системы

В системах второго типа человеку отводится роль собеседника, который может сам проявлять интерес к разным аспектам информации, задавая любые вопросы.

Диалог человека и машины

Общая блок-схема ориентированной на задачу системы речевого диалога



Диалог человека и машины

Классификация систем устного диалога по направлению потока информации

Тип системы	Поток информации	Примеры
Объяснение	Система → Пользователь	Прокладка маршрута, обучение
Заполнение форм	Пользователь → Система	Покупки по телефону, перенаправление абонента
Получение информации	Пользователь ↔ Система	Резервирование гостиницы, поиск литературы

Диалог человека и машины

Классификация систем устного диалога по структуре информации

Структура информации	Тип системы	Распознавание речи	Семантическая интерпретация	Стратегия диалога
Реляционная БД	Запрос к БД	Ограничен-ное грамматикой, заданной экспертом	Отображение результата понимания на SQL-запрос	Заполнение необходимых форм и подтверждение второстепенных
ЕЯ текст	Извлечение документа	Статистическая модель	Извлечение информации (векторно-пространственная	Уточнение запросов и ограничение совпадающих элементов