

Features in Speech Perception and Lexical Access

Kenneth N. Stevens

1. Цель работы – описать процессы, посредством которых человек извлекает последовательность лексических единиц из речевого потока. Слова хранятся в памяти человека в виде последовательностей сегментов, каждый из которых представляет собой набор бинарных смысловоразличительных признаков (*distinctive features*). Существует универсальный набор признаков, в каждом языке представлено подмножество признаков, играющих в нём смысловоразличительную роль.

2. Смысловоразличительные признаки

В основе смысловоразличительных признаков лежат физические свойства человеческого артикуляторного аппарата как генератора звука и свойства перцептивной системы, принимающей акустический сигнал.

Выделяются три типа признаков:

1) Артикуляторно-несвязанные признаки (*articulator-free features*) – базовые признаки, такие как: [гласный], [согласный], [глайд], а также более дробные признаки [сонорный], [сибилянт], которые делят согласные на смычные, свистяще-шипящие фрикативные, прочие фрикативные и сонорные.

2) Артикуляторно-связанные признаки (*articulator-bound features*) определяют активные органы речи и место образования.

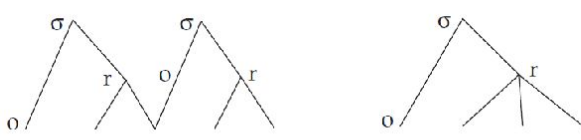
3) Дополнительные признаки (*additional features*) характеризуют непервичные артикуляторные свойства: назализация, лабиализация, звонкость.

Table 6.2 Listing of articulator-free features, articulators, and articulator-bound features for some consonants in English

	<i>b</i>	<i>d</i>	<i>g</i>	<i>p</i>	<i>f</i>	<i>s</i>	<i>z</i>	<i>ʃ</i>	<i>m</i>	<i>l</i>
continuant	–	–	–	–	+	+	+	+	–	–
sonorant	–	–	–	–					+	+
strident						+	+	+		
lips	+			+	+				+	
tongue blade		+				+	+	+		+
tongue body			+							
round	–			–	–				–	
anterior		+				+	+	–		+
lateral									–	+
high			+							
low			–							
back			+							
nasal									+	
stiff vocal folds	–	–	–	+	+	+	–	+		–

3. Лексическая репрезентация

Table 6.4 Lexical representations for the words *sudden* and *help* (The syllable structure of each word is schematized at the top: σ = syllable, o = onset, r = rime)



	<i>s</i>	<i>ʌ</i>	<i>d</i>	<i>ə</i>	<i>n</i>	<i>h</i>	<i>ɛ</i>	<i>l</i>	<i>p</i>
vowel		+		+			+		
glide						+			
consonant	+		+		+			+	+
stressed		+		–			+		
reducible		–		+			–		
continuant	+		–		–			–	–
sonorant			–		+			+	–
strident	+								
lips									+
tongue blade	+		+		+			+	
tongue body									
round									–
anterior	+		+		+			+	
lateral								+	
high		–		–			–		
low		–		–			–		
back		+		+			–		
tense		–					–		
spread glottis						+			
nasal					+				
stiff vocal folds	+		–						+

4. Вариативность акустических коррелятов смыслоразличительных признаков

На реализацию смыслоразличительных признаков могут влиять различные факторы, вызывая существенную вариативность:

- 1) просодический контекст: позиция относительно границ фразы и структуры слога, ударение;
- 2) дополнительные артикуляторные жесты, имеющие целью усилить перцептивный контраст некоторого признака;
- 3) наложение акустических жестов двух соседних сегментов, приводящее к ослаблению и даже утрате акустических коррелятов тех или иных признаков.

Несмотря на фонетическую вариативность, слушающий способен извлекать более или менее однозначно информацию о наборе смыслоразличительных признаков и идентифицировать цепочки слов.

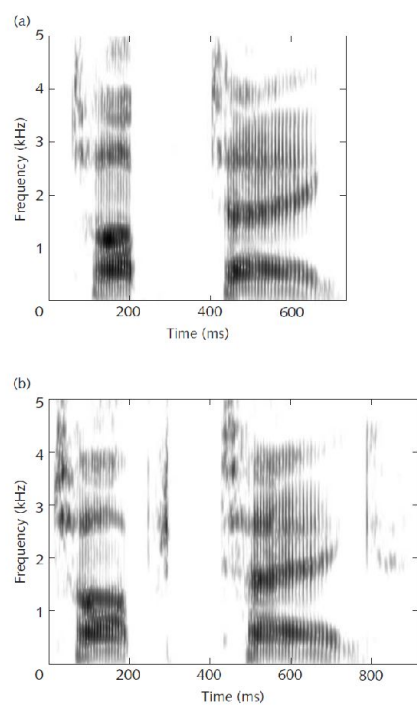


Figure 6.3 Spectrograms of two versions of *top tag* produced by a male speaker: (a) casually spoken version, in which the /p/ release and /t/ closure are not evident acoustically; (b) carefully spoken version in which these events can be seen in the spectrogram between about 240 and 300 ms.

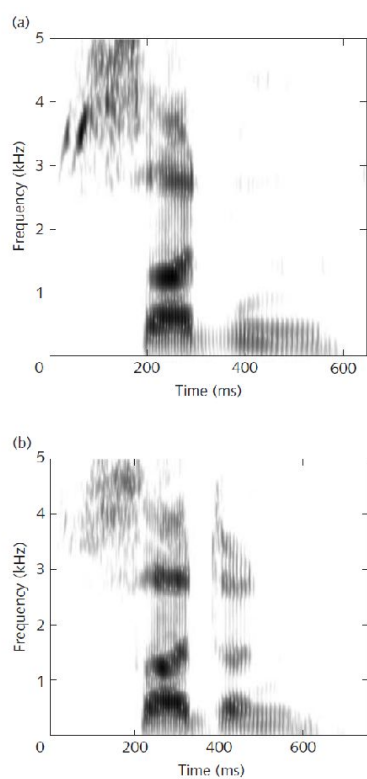


Figure 6.4 Spectrograms of the word *sudden* produced in a casual manner in (a) and a clearer manner in (b), produced by a male speaker. In (a) the /ʌn/ sequence is produced as syllabic nasal, whereas in (b) there is clear evidence for a vowel and a landmark at the time of /n/ closure, at about 480 ms.

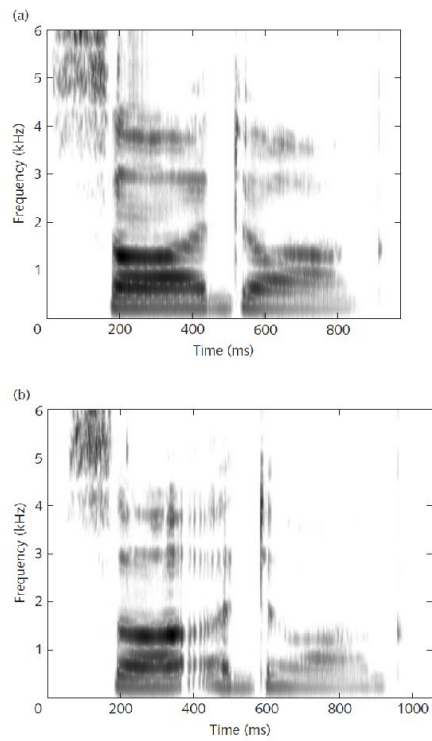


Figure 6.5 (a) Spectrogram of the utterance *saw a dog* produced in a casual manner. There is no separate amplitude peak for the weak vowel /ə/. (b) Spectrogram of the same utterance produced in a more careful way. The vowel /ə/ is separated from /a/ with glottalization, indicating a word boundary. In both utterances, indirect evidence for /ə/ appears in the time course of the F2 transition immediately preceding the /d/ closure. Female speaker.

5. Модель восприятия речи

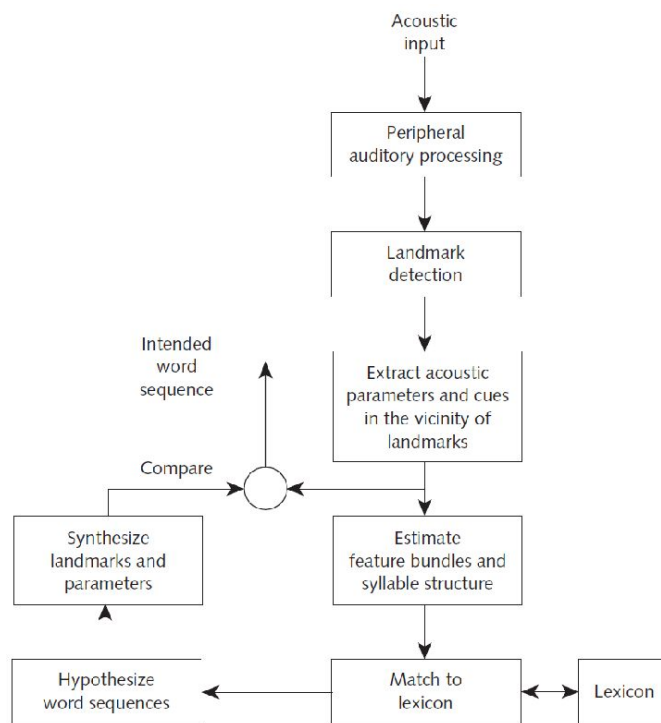


Figure 6.7 Block diagram of a model of human lexical access. The input at the top is the acoustic speech signal for an utterance produced by a speaker. The output is the sequence of words intended by the speaker. The model consists of a direct analysis path in which arrays of feature bundles are estimated and are matched to the lexicon to obtain hypothesized word sequences, and a synthesis path that determines whether a hypothesized word sequence could produce an acoustic pattern that matches the pattern derived from the acoustic signal.

- 1) Peripheral auditory processing: восприятие акустического сигнала, как речевого, так и неречевого.
- 2) Landmark identification: распознавание границ между сегментами на основе шаблонов изменения амплитудно-частотного спектра (усиление в низкочастотной области, скачкообразные изменения спектрального рисунка) и артикуляторно-несвязанных признаков.
- 3) Extraction of acoustic parameters and cues for features: уточнение артикуляторно-несвязанных признаков, распознавание артикуляторно-связанных признаков, распознавание структуры слога.
- 4) Estimating the distinctive feature bundles and syllable structure: распознавание смысловоразличительных признаков по следующей модульной схеме:

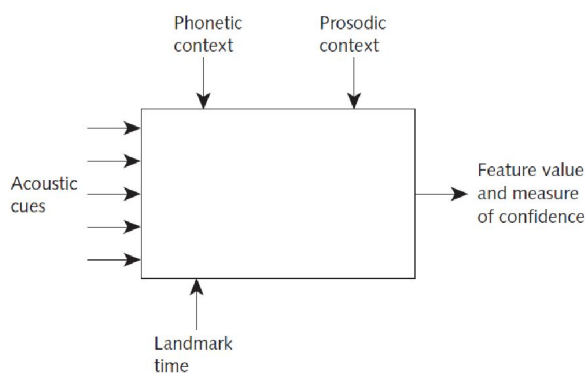


Figure 6.8 Schematic representation of a module for estimating an articulator-bound feature.

Модули для различных признаков должны быть связаны между собой.

5) Access to lexicon: поиск последовательности пучков признаков в лексиконе. Возможны три стратегии: хронологическая (от первого сегмента к последнему), вероятностная (учитывающая помехи и ошибки), структурная (от вокалического ядра слога к остальным элементам слога).

6) Verifying hypothesized words by synthesis: синтез идентифицированных пучков признаков, сверка результата синтеза с полученным на вход акустическим сигналом и выбор оптимального варианта.