

УДК 621.391:681.3.06

ЭФФЕКТ МАСКИРОВКИ И АВТОМАТИЧЕСКИЙ АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ

Н.Г.Загоруйко, В.Г.Лебедев

Одной из важнейших проблем, стоящих в настоящее время перед исследователями, занимающимися автоматическим распознаванием речи, является проблема выбора эффективной системы признаков для описания речевого сигнала. Для решения этого вопроса вполне естественно желание наиболее полно использовать имеющуюся информацию о механизмах анализа речи, присущих слуховой системе человека.

Существуют три основные гипотезы слухового описания сигналов со сложным спектром: гипотеза полного спектрального описания, полосного описания и форматного описания [1]. Соответственно существуют три метода описания сигнала при автоматическом распознавании речи.

Согласно первому методу для фонемного решения используется полная спектральная картина сигнала без какой-либо предварительной обработки. Точно так же описываются и эталоны.

Полосный и формантный методы предполагают процедуру выделения из полного спектрального описания некоторых признаков, что позволяет получить более простое и экономичное описание по сравнению с полным. В полосном методе в качестве признаков используются значения энергий в широких частотных полосах, например, 300-600 гц, 600-1200 гц, 1200-2400 гц и другие.

Сокращение описания в формантном методе достигается за счет выделения из полного спектрального описания сигнала значений частот формантных максимумов, которые и используются в качестве признаков для принятия фонемного решения.

В последнее время появились работы по экспериментальному исследованию восприятия синтетических речевых стимулов, позволяющие опровергнуть гипотезы полного и полосного слухового описания и подтверждающие форматную гипотезу [2]. Данные о том, что для принятия фонемного решения человек использует не весь спектр, а лишь его локальные признаки (особенности) были получены в целом ряде экспериментов, проведенных различными исследователями [3-5].

В настоящей работе в качестве механизма для выбора признаков из спектрального описания речевого сигнала предлагается использовать машинную модель "эффекта маскировки", присущего слуховой системе человека.

В психоакустике известны два метода измерения кривых одновременной маскировки одних сигналов (обычно тональных) другими сигналами [6].

В первом методе маскирующий сигнал (маскер) имеет постоянную частоту и амплитуду, а на различных частотах измеряется пороговая амплитуда испытательного тона, т.е. та амплитуда, при которой испытательный тон перестает восприниматься.

Во втором методе, наоборот, испытательный тон имеет постоянную частоту и амплитуду, а на различных частотах измеряется амплитуда маскера, при которой испытательный тон перестает восприниматься.

Наклон кривых маскировки в сторону низких частот, измеренный в экспериментах по первому методу (и в сторону высоких частот по второму методу) на частотах выше 500 гц составляет величину порядка 90-280 дб на октаву, а в сторону высоких частот (низких по второму методу) - 5-40 дб на октаву [7],[8].

Имеется также много экспериментальных результатов исследования маскировки во временной области [9-12]. Здесь наблюдаются эффекты прямой и обратной маскировок.

Кривая прямой маскировки выражает зависимость порога восприятия испытательного тона, следующего через некоторый временной интервал после маскера, от величины этого интервала. Эф-

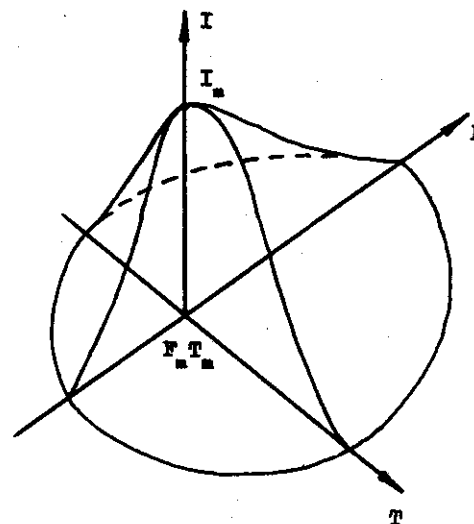


Рис. 1

Если представить маскер как тональный сигнал интенсивности I_m в трехмерном пространстве с измерениями F, T, I , где F - частота, T - время, I - интенсивность, то его маскирующее влияние геометрически можно изобразить в ви-

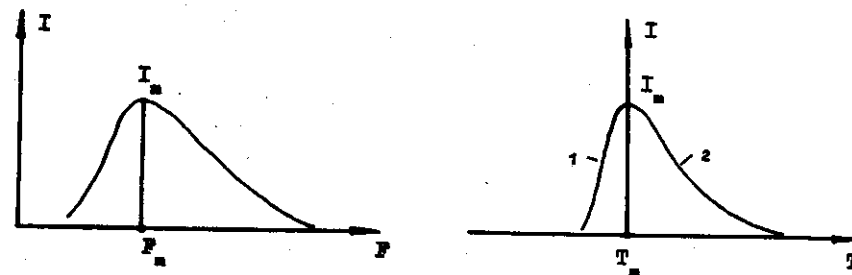


Рис. 2

де поверхности, имеющей вид несимметричного "колокольчика", изображенного на рис. 1 (I_m - интенсивность маскера; F_m, T_m - частотная и временная координаты маскера). Сечение этой поверх-

фективный интервал прямой маскировки (расстояние между маскером и тоном, на котором этот эффект наблюдается) составляет 150-250 мсек.

Кривая обратной маскировки выражает зависимость порога восприятия испытательного тона, предшествующего маскеру от величины временного интервала между тоном и маскером. Эффективный интервал обратной маскировки равен 20-30 мсек.

Если представить маскер как тональный

ности плоскостью F , T_m , T есть кривая одновременной маскировки (рис. 2, а), а сечение плоскостью F_m , T , T - кривая, состоящая из ветвей обратной и прямой маскировок (ветви 1 и 2-на рис. 2, б).

Основная идея предлагаемого метода выделения признаков состоит в том, чтобы, построив на каждой энергетической составляющей спектрально-временной картины речевого сигнала "колокольчик" маскировки, удалить те составляющие, которые будут накрыты (замаскированы) соседними "колокольчиками".

Для машинной реализации метода была построена математическая модель поверхности маскировки. В качестве аппроксимирующего выражения кривой одновременной маскировки принято следующее [15]:

$$T_j = T_m \cdot \exp \left[-1,44 \cdot Q^2 \left(\frac{\omega_m}{\omega_j} - 1 \right)^2 \right], \quad (1)$$

где T_m - интенсивность маскера, ω_m - частота маскера, T_j - пороговая интенсивность тона, ω_j - частота тона, $Q = 1 - \frac{1}{5}$.

Кривая прямой маскировки аппроксимируется [14] выражением вида:

$$T_j = \beta \cdot T_m \left\{ 1 - \alpha \cdot \left[1 - \exp \left(-\frac{k \cdot \Delta t}{\tau} \right) \right] \right\}, \quad (2)$$

где $k = j - m$, $\alpha = 0,7 \div 0,8$, $\tau = 50$ мсек, $\beta = 1$, а для кривой обратной маскировки было принято выражение:

$$T_j = \beta \cdot T_m \cdot \exp \left(-\frac{k \cdot \Delta t}{\tau} \right), \quad (3)$$

где $\tau = 25$ мсек, $k = j - m$, $\beta = 1$.

Точки поверхности маскировки определяются путем совместного решения системы из двух уравнений (уравнения 1 и 2 для значений $T > T_m$ и уравнений 1 и 3 для значений $T < T_m$).

Ниже описывается алгоритм "колокольчик", разработанный для выделения спектральных параметров речевого сигнала, моделирующий эффект маскировки.

Исходный акустический сигнал через аналого-цифровой преобразователь с частотой квантования 20 кГц вводится в ЭВМ "БЭСМ-6" и записывается на магнитную ленту.

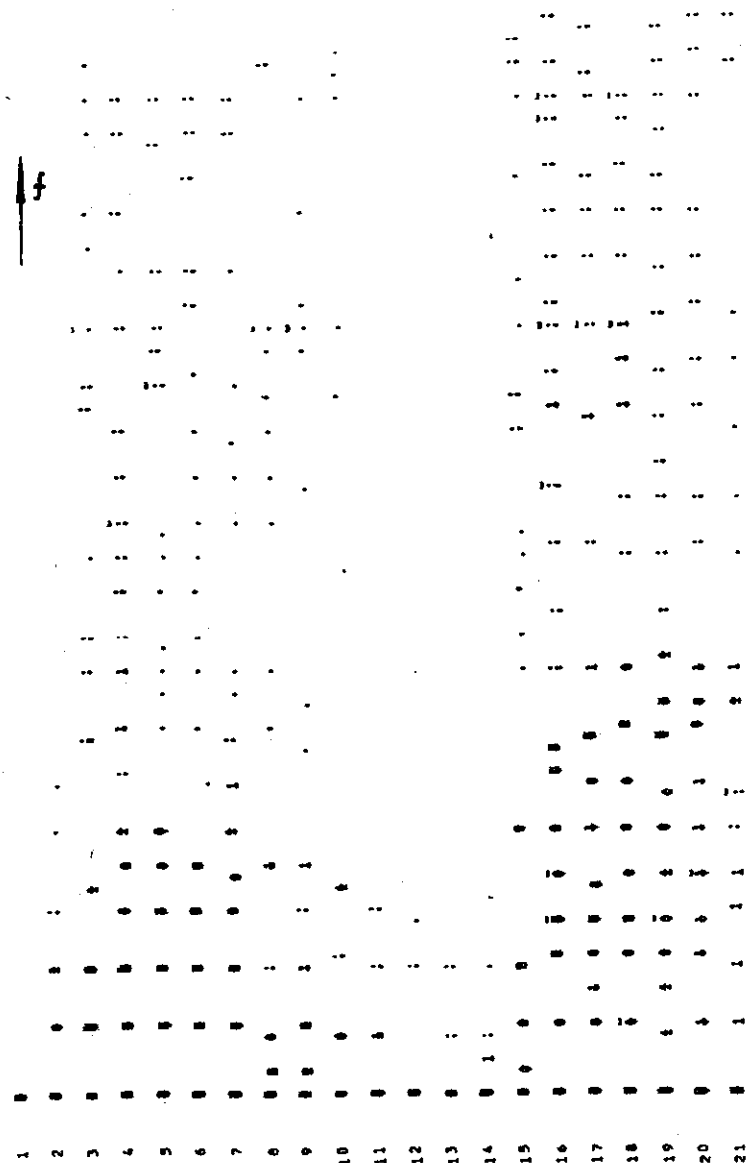


Рис. 3. Исходная спектрограмма слова "полна".

f

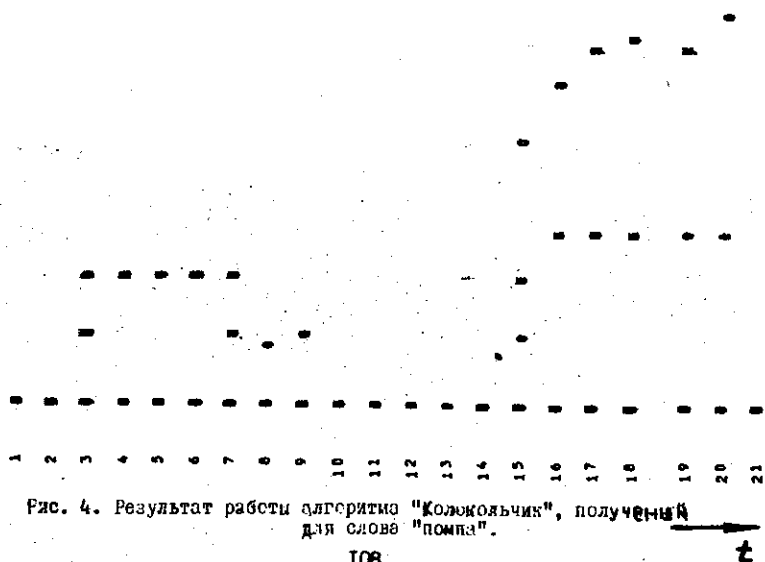


Рис. 4. Результат работы алгоритма "Колокольчик", полученный для слова "pompa".

Анализируемое слово разбивается на сегменты фиксированной длительности $T = \Delta t$ мсек (ℓ отсчетов). Величина Δt задается предварительно и может меняться в пределах от 1 до 25 мсек. Внутри каждого сегмента мгновенные значения сигнала взвешиваются и вычисляются значения C_i ($i = 1, 2, \dots, \rho$) модулей спектра по Фурье (здесь $\rho = \frac{\ell}{2} + 1$). Для каждого значения C_i строится частотная группа [13]. Ширина частотных групп в области частот ниже 500 гц составляет примерно 100 гц, а в области выше 500 гц она увеличивается пропорционально средней частоте, т.е.

$$\Delta f \approx 0,2 f_c,$$

где Δf - ширина частотной группы, f_c - средняя частота.

Внутри каждой (i -й) группы вычисляются значения энергии E_i

$$E_i = \sum_{j=i-\frac{\Delta f}{2}}^{i+\frac{\Delta f}{2}} \quad (i = 1, 2, \dots, \rho), \quad (4)$$

где Δf_i - ширина i -й частотной группы (в отсчетах). Затем выделяются все локальные максимумы энергий E'_i и для них проверяется условие взаимной маскировки по частоте

$$E'_i \geq E'_j \exp \left[-1,44 Q^2 \left(\frac{\omega_j}{\omega_i} - 1 \right)^2 \right], \quad (5)$$

где ω_i - средняя частота i -й группы ($i = 1, 2, \dots, \rho, j = 1, 2, \dots, \rho$). Замаскированные значения энергий удаляются из дальнейшего рассмотрения.

Следующий этап работы алгоритма - маскировка во временной области. Для всех оставшихся значений E'_j каждого сегмента проверяется условие прямой маскировки на глубину 100-200 мсек.

$$E'_{i/2} < \bar{E}'_{i/1} \cdot \beta \cdot \left\{ 1 - \alpha \left[1 - \exp \left(-\frac{k \Delta t}{\tau} \right) \right] \right\}, \quad (6)$$

где $k = i - n$, $\tau = 50$ мсек, $\alpha = 0,7 - 0,8$, $\beta = 1$.

Затем проверяется условие обратной маскировки на глубину 25 мсек.

$$E'_{ij_2} < \bar{E}'_{nj_1} \cdot \beta \cdot \exp \left(- \frac{k \Delta t}{\tau} \right),$$

где $\tau = 25$ мсек,

$$\bar{E}'_{nj_1} = E'_{nj_1} \cdot \exp \left[-1,44 Q^2 \left(\frac{\omega_{j_1}}{\omega_{j_2}} - 1 \right)^2 \right],$$

$$j_1 = 1, 2, \dots, p, \quad j_2 = 1, 2, \dots, p.$$

n - номер маскирующего сегмента,

i - номер маскируемого сегмента:

$i = n+1, n+2, \dots, n+G$ (для прямой маскировки),

$i = n-1, n-2, \dots, n-G$ (для обратной маскировки),

G - глубина маскировки.

Незамаскировавшиеся значения E'_i в каждом сегменте принимаются за информативные признаки этого сегмента (назовем их маскипризнаками). Описанный алгоритм реализован в виде комплекса подпрограмм на языке "Фортран" для ЭВМ "БЭСМ-6".

Проведены эксперименты по выделению маскипризнаков речевого сигнала. На рис.3 и 4 приведены результаты работы алгоритма "колокольчик", полученные для слова "помпа".

Предварительная оценка полученных для различных речевых сигналов спектрограмм показывает, что число выделяемых маскипризнаков колеблется от 3 до 10 на сегмент, а положение признаков на частотно - временных картинках примерно соответствует формантной структуре анализируемых речевых сигналов.

Возможны два метода оценки качества полученных признаков:

1. Обратный синтез речевого сигнала по маскипризнакам и сравнение его по качеству звучания с сигналом, синтезированным по полному спектральному описанию.

2. Оценки по результатам распознавания.

Первый метод кажется наиболее простым, но, к сожалению, мы не располагаем необходимой для синтеза аппаратурой, поэтому был выбран второй метод.

В настоящее время проводятся эксперименты по автоматическому распознаванию речевых сигналов с использованием маскипризнаков.

В заключение авторы выражают благодарность В.М.Величко за ценные обсуждения и В.С.Лозовскому за любезно предоставленную возможность воспользоваться некоторыми его подпрограммами.

Л и т е р а т у р а

1. ЧИСТОВИЧ Л.А., КОЖЕВНИКОВ В.А. Восприятие речи. - В кн.: Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1969, с. 4-150.

2. МУШНИКОВ В.Н. Экспериментальное исследование слуховой обработки стационарных синтетических гласных. Автореф. диссер. Л., 1973, 23 с. Ин-т физиологии им. Н.П.Павлова, АН СССР.

3. ШУПЛЯКОВ В.С. О тональной высоте звуков (S) и (S̄). - В кн.: Механизмы речеобразования и восприятия сложных звуков. М.-Л., 1966, с. 87-95.

4. LINDQVIST I.Ch.K.-G., PAULI S. The role of relative spectrum levels in vowel perception. - In: Reports of the 6th Intern. Congr. on Acoustics, Tokyo, 1968, vol.2:B, p.91-94.

5. МУШНИКОВ В.Н., ЧИСТОВИЧ Л.А. О слуховом описании гласного. - В кн.: Анализ речевых сигналов человеком. Л., Изд-во "Наука", 1971, с.5-19.

6. ШУПЛЯКОВ В.С. Аппроксимация кривых одновременной маскировки слуха. - В кн.: Вопросы теории и методов исследования восприятия речевых сигналов, № 2, Л., 1971, с. 5-20.

7. ЧИСТОВИЧ Л.А. Частотные характеристики эффекта маскировки. - "Биофизика", 1957, т.2, с.749-755.

8. SMALL A.M. Pure-Tone Masking. - "J.Acoust.Soc.of Amer.", 1959, vol.31, N 12, p.1619-1625.

9. САМОЙЛОВА Н.К. Зависимость величины маскировки коротких сигналов от интервала между звуками. - "Биофизика", 1959, т.4, вып.У, с. 550-558.

10. ЧИСТОВИЧ Л.А., ИВАНОВА В.А. Взаимная маскировка коротких звуковых импульсов. - "Биофизика", 1959, т.4, вып.2, с.170-180.

11. PICKETT J.M. Backward Masking. - "J.Acoust. Soc. of Amer.", 1959, vol.31, N 12, p.1613-1615.

12. RAAB D.H. Forward and Backward Masking between Acoustic Clicks. - "J.Acoust.Soc.of Amer.", 1961, vol.33, N 1, p.137.

13. ПИКЕТТ Э., ФЕЛЬДКЕЛЛЕР Р. Ухо, как приемник информации. М., Изд-во "Связь", 1971.

14. МУШНИКОВ В.Н., ЧИСТОВИЧ Л.А. Проявление адаптации при восприятии речевого сигнала. - В кн.: Вопросы теории и методов исследования восприятия речевых сигналов, № 2, Л., 1971, с.28-35.

15. БОНДАРЕНКО В.П. Разработка и исследование электронной модели периферии слуховой системы человека. Автореф. диссертации, на соискание ученой степени канд.техн.наук. Томск, 1972, 20 с. Томский политехн. ин-т им.С.М.Кирова.

Поступила в ред.-изд.отд.
17 января 1974 года