

Остальные члены описывают спектры высших порядков, возникающие из-за нелинейности процесса модуляции. При балансной модуляции в дифракционной картине отсутствуют спектры четных порядков, в том числе и спектр нулевого порядка. Устранение этого спектра позволяет визуализировать звуковые волны, приходящие по нормали к решетке приемников. На фиг. 3 приведены фотографии дифракционных картин, снятые в режиме балансной модуляции при углах падения $\theta=0, 20, 40, 60$. Видно образование одной светлой полосы при нормальном падении (а) и двух симметричных полос, смещения которых от центра дифракционной картины увеличивается с возрастанием угла падения (б, в, г). Величины этих смещений, найденные путем фотоэлектрических измерений, практически точно совпадают с рассчитанными теоретически по формуле (4).

Авторы благодарят В. М. Панкратова и Н. Н. Фомичева за предоставленный им модулятор света.

ЛИТЕРАТУРА

1. Шендеров Е. Л. Формирование звуковых изображений в фазовой голографической системе. В кн.: Радио- и акустическая голография. Л., «Наука», 1976, 119.
2. Свет В. Д. Методы акустической голографии. ЦНИИ «Румб», 1976.

Поступила
21 февраля 1978 г.

УДК 534.782

ОПРЕДЕЛЕНИЕ СЕГМЕНТОВ РАВНЫХ АРТИКУЛЯЦИЙ ПО РЕЧЕВОМУ СИГНАЛУ

А. С. Рылов

Передача данных в слуховую оперативную память осуществляется специальным механизмом выделения моментов резких изменений в сигнале [1, 2], обусловленных перестройкой органов артикуляции. В соответствии с этим подход к проблеме сегментации может быть сведен к определению некоторого критерия сходства между текущими функциями площади сечений речевого тракта. Таким критерием может быть коэффициент корреляции [3]

$$(1) \quad r_i = \frac{\sum_{m=1}^M (s_m - M_s)(s'_m - M_{s'})}{\sqrt{\sum_{m=1}^M (s_m - M_s)^2 \sum_{m=1}^M (s'_m - M_{s'})^2}}$$

где M_s и $M_{s'}$ — математические ожидания векторов s и s' (ФПС) в t и $t+1$ кадрах анализа речевого сигнала, а s_m и s'_m — координаты векторов s и s' (площади сечений), для определения которых речевой тракт от связок до губ представляется отрезками труб длиной l . При этом предполагается:

1) вся труба жесткая и потери звуковой волны вследствие вязкости и температурных изменений несущественны;

2) поперечный размер каждого сечения мал по сравнению с длиной волны (волна плоская);

3) диаметры сечений s_m и s_{m+1} отличаются незначительно.

Тогда для определения s_m можно воспользоваться известной формулой [4]

$$s_m = (1 + \mu_m) s_{m+1} / (1 - \mu_m),$$

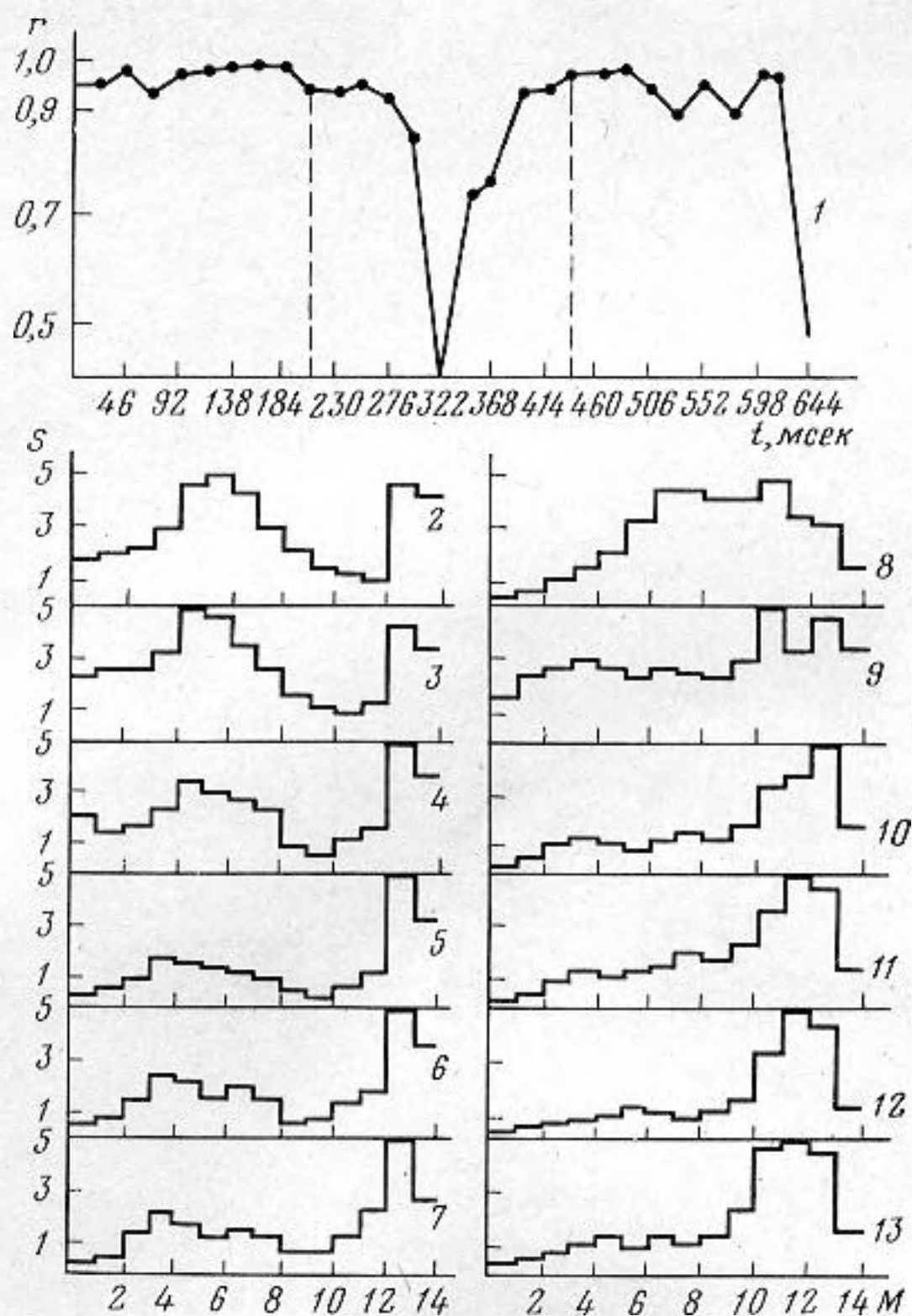
где μ_m — коэффициенты отражения на границе двух сечений. Кроме того, известно [5], что процессы фильтрации в обратном регрессивном фильтре и в трубе, описанной выше, при некоторых дополнительных условиях аналогичны, и можно положить

$$\mu_m = k_{m-1},$$

где k_m — коэффициенты частичной корреляции в обратном регрессивном фильтре [6]. Таким образом, векторы s и s' определяются непосредственно из речевого сигнала и подставляются в формулу (1) для получения сегментирующих функций.

Чтобы выделить переходные и стационарные участки (сегменты равных артикуляций) в сегментирующих функциях, используется алгоритм, анализирующий пять условий: 1) $p \leq r_i \leq r_{i+1}$, 2) $p \leq r_{i+1} \leq r_i$, 3) $r_{i+1} \leq r_i < p$, 4) $r_i \leq r_{i+1} < p$, 5) $r_i < p \leq r_{i+1}$; здесь r_i — коэффициент корреляции в сегментирующей функции, p — пороговая константа.

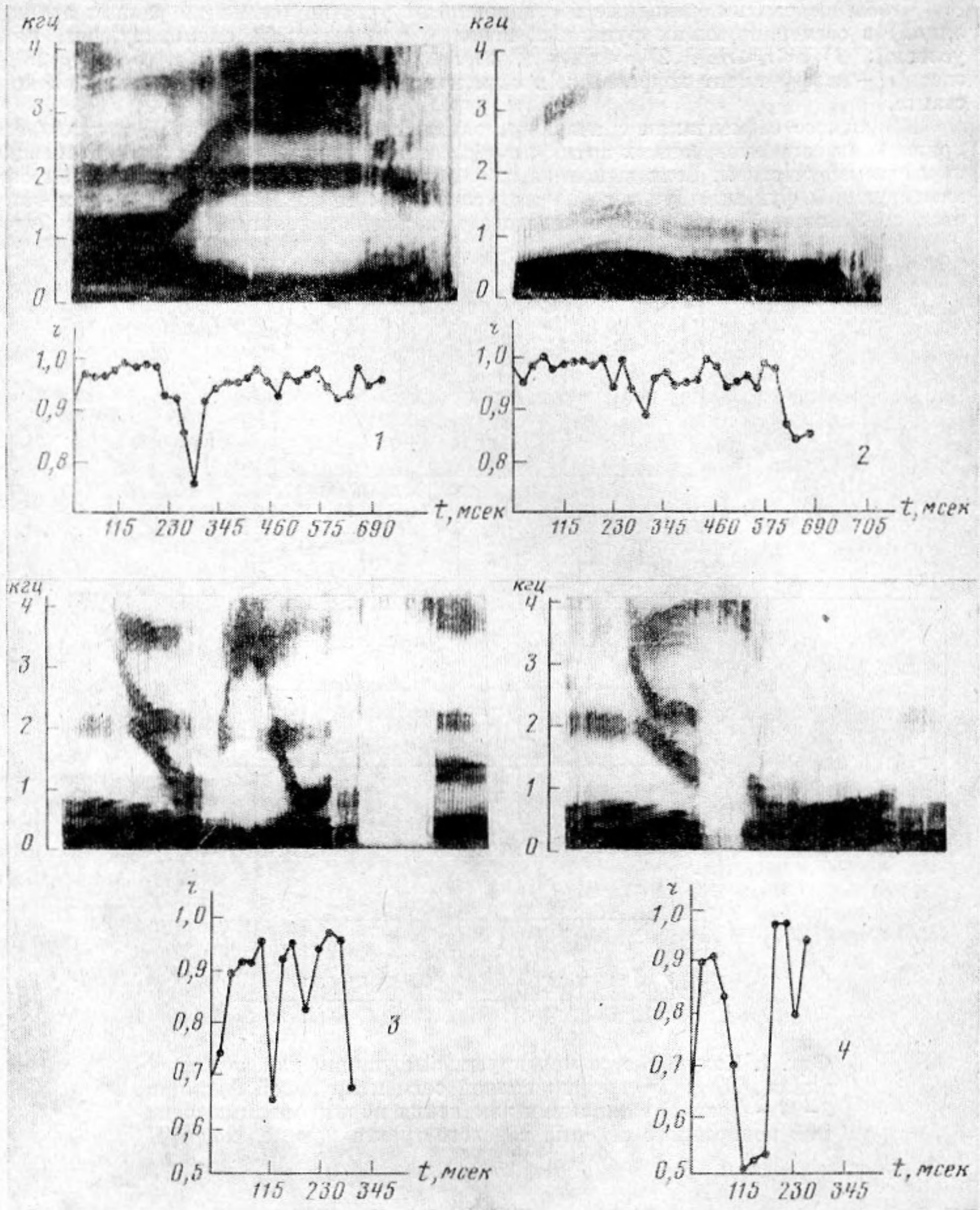
В процессе сегментации сначала выполняются условия 1) или 2), являющиеся признаками сегментов равных артикуляций, затем условия 3) и 4), соответствующие переходным участкам. Если выполняется условие 3), то локальным минимумом сегментирующей функции будет r_{i+1} , если условие 4), то локальным минимумом останется r_i . Условия 3) и 4) могут чередоваться, тогда локальным минимумом будет



Фиг. 1. Получение сегментирующей функции для звукосочетания /ИА/: 1 — график полной сегментирующей функции; 2—13 — фрагмент динамики изменения нормированных функций поперечного сечения речевого тракта при переходе /И/ в /А/

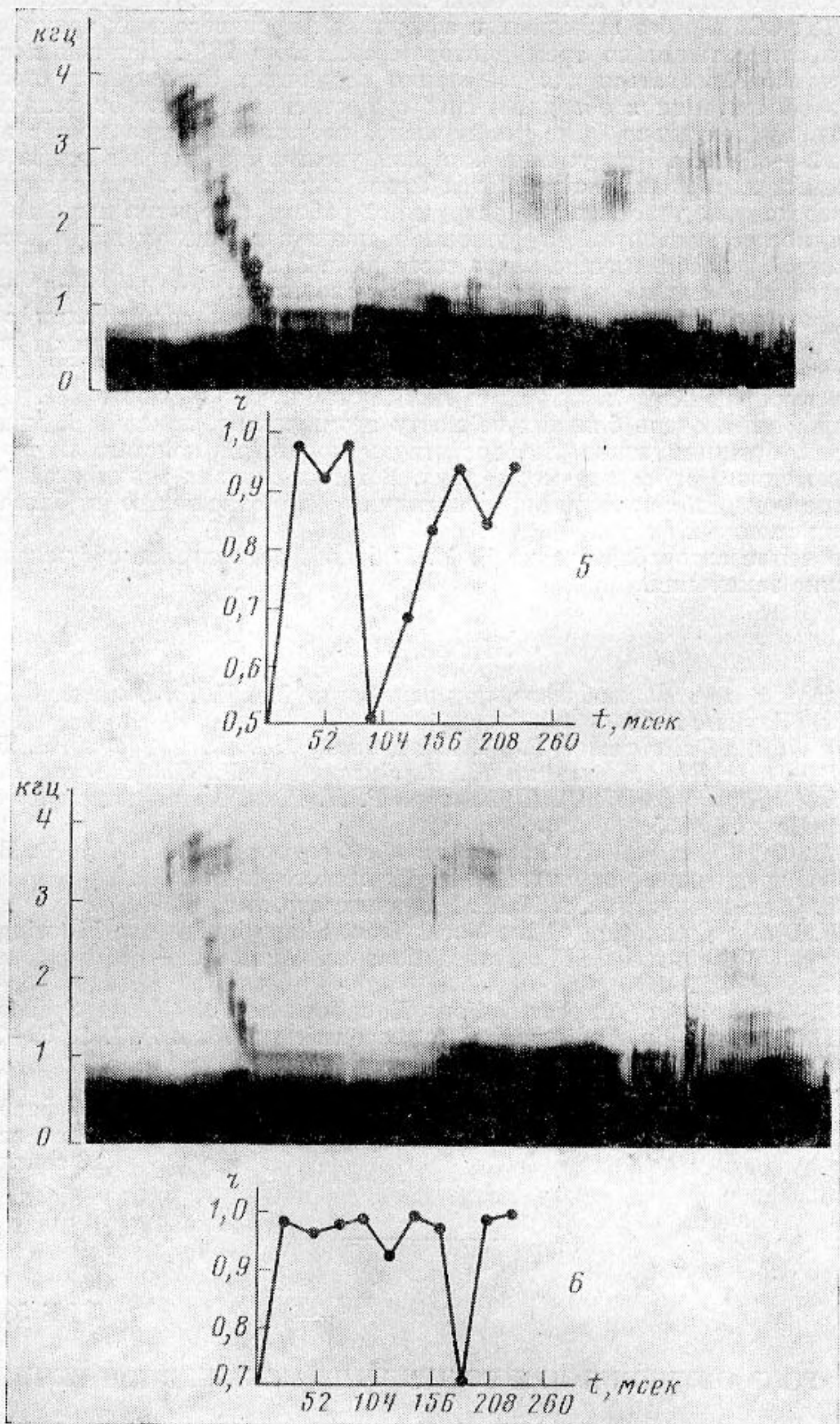
минимальное значение из всех r_{i+1} для условия 3). И, наконец, условие 5) показывает, что закончился переходный участок, а локальным минимумом сегментирующей функции должно остаться последнее минимальное значение. Затем этот цикл повторяется для нового стационарного и нового переходного участков. Таким образом, сегменты равных артикуляций будут находиться между двумя локальными минимумами, определенными по описанному выше алгоритму.

Для экспериментальной проверки метода сегментации записывались образцы звукосочетаний типа ГГ, СС, СГ от трех дикторов (двое мужчин и одна женщина). Запись производилась через АРУ с нормированием уровня сигнала. После чего были сняты сонограммы на звуковом спектрографе типа В/65 фирмы КЭИ. Сонограммы снимались в диапазоне 8 кгц для дикторов-мужчин и 16 кгц для диктора-женщины, поэтому на сонограммах соответственно в одном сантиметре 115 и 52 мсек. Затем речевой материал вводился в ЭВМ, предварительно проходя через фильтр с частотой среза 5,4 кгц, в связи с чем частота дискретизации в АЦП была выбрана 11 кгц. После этого речевой сигнал обрабатывался по алгоритму, описанному выше, имеющему на выходе сегментирующую функцию (фиг. 1, кривая 1). На фигуре изображен также фрагмент динамики изменения нормированных функций площади сечения речевого тракта для звукосочетания /ИА/, охватывающий временной интервал, в который входит конец стационарного участка звука /И/ (кривые 2—6), переходной участок от /И/ до /А/ (кривые 7—10) и начало стационарного участка звука



Фиг. 2. Сонограммы и сегментирующие функции звукосочетаний: 1 — /АИ/; 2 — /ОУ/; 3 — /ВЬЮ/ в слове НЕ ВЫЮТСЯ; 4 — /ЗВ/ в слове НЕ ЗВОН, диктор-мужчина, 5 — /МЛ/ в слове НЕ МЛЫН, 6 — /МН/ в слове НЕ МНОГО, диктор-женщина

/А/ (кривые 11–13). На кривой 1 этот фрагмент отмечен штриховыми линиями. На данной кривой имеются нестационарные участки, локальные минимумы которых выше пороговой величины, равной в данном случае 0,9. Поэтому они относятся к стационарной части фонемы. Кроме того, минимум изменения корреляции достигается не одномоментно, а растянут на четыре кадра, отражая постепенно нарастающее усилие при изменении положения артикуляторных органов. После определения сегментирующей функции производилось сравнение сегментов равных артикуляций на ней с длительностями стационарных участков фонем на сонограмме (фиг. 2). На данной фигуре кривые 1 и 2 являются сегментирующими функциями звукосочетаний /АИ/ и /ОУ/ для диктора-мужчины. Локальные минимумы этих функций приходятся на середину переходного участка между фонемами на соответственных сонограммах, расположенных выше. Локальный минимум на кривой 1 выражен более отчетливо, чем на кривой 2. Это связано с тем, что фонемы /А/ и



Фиг. 2

/И/ значительно отличаются по месту артикуляции по сравнению с фонемами /О/ и /У/. Кривые 3 и 4 являются сегментирующими функциями звукосочетаний /ВЬЮ/ и /ЗВ/ в словах НЕ ВЬЮТСЯ, НЕ ЗВОН для диктора-мужчины, сонограммы которых помещены над кривыми. Начала координат кривых 3 и 4 располагаются в началах первых фонем каждого звукосочетания. Здесь также видно четкое отделение фонем. Кроме того, на кривой 3 второй локальный минимум разделяет дифтонг, состоящий из короткого /И/ и длинного /У/. Кривые 5 и 6 являются сегментирующими функциями звукосочетаний /МЛ/ и /МН/ в словах НЕ МЛЫН, НЕ МНОГО для диктора-женщины. На сонограммах этих слов границы между фонемами звукосочетаний /МЛ/ и /МН/ выражены значительно хуже, чем на сегментирующих функциях. Локальный минимум звукосочетаний /МН/ выражен слабо, так как /М/ и /Н/ очень близки по месту артикуляции, однако визуальное членение по сегментирующей функции можно произвести легче, чем по сонограмме.

С помощью описанного метода была произведена сегментация около 150 звуко-сочетаний ГГ, СС, СГ (без взрывных и аффрикат), произнесенных тремя дикторами. Надежность сегментации по трем дикторам составила 98%. Данный метод, как и в [3], может использоваться для выделения первичных сегментов в слове, причем надежность сегментации в обоих случаях существенно зависит от величины кадра анализа. Экспериментально было установлено, что эта величина может выбираться в пределах 20–30 мсек. При уменьшении или увеличении кадра анализа от указанных временных интервалов наблюдается менее выраженная декорреляция векторов s и s' на переходных участках, что затрудняет работу алгоритма выделения локальных минимумов сегментирующей функции и понижает в целом надежность сегментации (в данной работе величина кадра составляла 23 мсек).

Сравнительный анализ двух методов (с использованием функций поперечного сечения и спектральных срезов [3]) показал, что время сегментации с помощью описанного метода сокращается в 7–8 раз, так как число координат в векторах s и s' составляет всего 12–14 вместо 100. Кроме того, переходные участки отмечаются более выраженными локальными минимумами, чем в работе [3]. Однако при сегментации фонем, очень близких по месту артикуляции, локальный минимум сегментирующей функции может не достигнуть пороговой константы, что приведет к выделению одного сегмента вместо двух. В таких случаях могут оказаться полезными другие методы сегментации, не критичные по отношению к близости места артикуляции смежных фонем.

В заключение автор благодарит Б. М. Лобанова за полезное обсуждение работы и критические замечания.

ЛИТЕРАТУРА

1. Распознавание слуховых образов, под ред. Загоруйко Н. Г., Волошина Г. Я. Новосибирск, «Наука», 1970.
2. Чистович Л. А., Кожевников В. А., Алякринский В. В., Бондарко Л. В., Голузина А. Г., Клаас Ю. А., Кузьмин Ю. И., Лисенко Д. М., Люблинская В. В., Федорова Н. А., Шупляков В. С., Шуплякова Р. М. Речь. Артикуляция и восприятие. М.—Л., «Наука», 1965.
3. Деркач М. Ф., Гумецкий Р. Я., Мишин Л., Оверченко М., Чабан М. Восприятие речи в распознающих моделях. Изд-во Львовского ун-та, 1971.
4. Скучик Е. Основы акустики, т. II. М., Изд-во иностр. лит., 1959.
5. Wakita H. Direkt Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, 5, 417–427.
6. Markel J., Gray A. On Autocorrelations Equations as Applied to Speech Analysis. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, 2, 69–79.

Минский радиотехнический институт

Поступила
17 февраля 1978 г.
После повторного исправления
26 сентября 1978 г.

УДК 534.684:534.42

РОЛЬ ОБУЧЕНИЯ ПРИ ВОСПРИЯТИИ ИСКАЖЕННОЙ РЕЧИ

М. А. Сапожков

Приведены данные по распознаванию речи слухом в случае амплитудно-частотных искажений ее. Определена роль фактора тренировки к искажениям речи и показано, что при длительной тренировке и обучении слушатель может хорошо воспринимать сильно искаженную речь.

Несмотря на ряд работ [1], до сих пор некоторые исследователи сомневаются, что формантные картины речи (взаимное расположение формант и антиформант и форма формантных переходов) являются определяющим признаком для распознавания звуков речи слухом человека [2]. Чтобы рассеять эти сомнения, были проведены эксперименты и было установлено, что при сильном подчеркивании спектральных составляющих речи около 300 гц звук «о» воспринимается как звук «у», а при подчеркивании их около 500 гц звук «у» воспринимается как звук «о», при подчеркивании спектральных составляющих в области 7000 гц звук «ш» воспринимается как звук «с», а при подчеркивании их в области около 4000 гц звук «с» воспринимается как звук «ш». При этих экспериментах слушатели не знали, что передают, и могли узнавать передаваемые слова или бессмысленные слоги или по смыслу (для