

МЕТОД ИССЛЕДОВАНИЯ РЕШАЮЩИХ ПРАВИЛ, ПРИМЕНЯЕМЫХ ПРИ ВОСПРИЯТИИ РЕЧИ

Л. А. ЧИСТОВИЧ*

Задача исследования распознавания фонем человеком состоит, в основном, в установлении акустических параметров речевого сигнала, измеряемых слухом, и в выявлении решающих правил, на основании которых звук определяется как та или иная фонема.

Первым шагом к решению этой задачи является определение и строгое количественное описание фонемных категорий. Под фонемной категорией понимается множество стимулов, воспринимаемых как одна и та же фонема. Акустический сигнал удобно рассматривать как точку в многомерном пространстве акустических параметров. Тогда фонемная категория будет соответствовать области в этом многомерном пространстве.

Очевидно, что для того, чтобы описать фонемную область, нужно указать, где находятся ее границы. Если граница между некоторыми двумя фонемами есть функция нескольких акустических параметров, она может иметь вид гиперповерхности. Логично предположить, что одна такая гиперповерхность будет разделять не только две частные фонемы, но две группы фонем, например, глухие и звонкие или мягкие и твердые. Тогда эта гиперповерхность будет отражать решающее правило, позволяющее перейти от акустических параметров к дифференциальному признаку: если сигнал находится по одну сторону гиперповерхности, ему присваивается одно значение по дифференциальному признаку, если он находится по другую сторону — другое значение. Мне кажется, что такое понимание дифференциального признака как решающей границы, разделяющей два класса фонемных категорий в пространстве акустических параметров — весьма близко к тому, которое дал Г. Фант в своей недавней статье „Природа дифференциальных признаков“ (1).

Итак, задача состоит в том, чтобы экспериментально определить положение границ между фонемными категориями в пространстве акустических параметров.

В принципе эту задачу можно решать обычным методом исследования фонемной идентификации синтетических речевых стимулов. Так она

* Институт физиологии им. И. П. Павлова АН СССР, Ленинград.

обычно и решается, если определяется граница между двумя фонемами по одному акустическому параметру. Однако, уже для трехмерного случая оказывается необходимым заранее записать на пленку и предъявлять для прослушивания так много различных стимулов, что работа становится очень трудно выполнимой. Дополнительная сложность состоит в том, что результаты идентификации стимулов разными испытуемыми нельзя объединять, т. к. положение границ у разных людей может не совпадать. Отсюда следует, что необходимо найти какой-то более быстрый и экономный метод определения границ между фонемными категориями.

В моем докладе я хочу рассказать об одном таком более быстром методе и привести, в качестве иллюстрации, некоторые результаты, полученные с его помощью.

Этот метод может быть назван методом активного поиска фонемной границы. Он является вариантом широко известного в психоакустике метода установки. Суть его состоит в том, что испытуемый сам управляет синтезатором сигналов, изменяя характер стимула, который он слушает. Испытуемому дается задание вращать только одну ручку прибора. Повернув ручку, испытуемый должен затем включить сигнал, прослушать его, снова повернуть ручку и продолжать это до тех пор, пока не удастся найти точку перехода от одной фонемы к другой. В простейшем случае ручка управления может быть связана только с одним параметром стимула. Остальные параметры стимула устанавливаются экспериментатором и остаются фиксированными во время поиска границы. Можно также одной ручкой управления изменять одновременно несколько параметров стимула, для чего экспериментатор должен задать между этими параметрами определенную функциональную зависимость. Поворот ручки будет соответствовать перемещению точки, представляющей сигнал, по некоторой траектории в пространстве акустических параметров.

На рис. 1 влево приведены для примера две траектории в пространстве двух параметров. Одна из них соответствует случаю, когда ручка связана только с одним параметром стимула, значение стимула по второму пара-

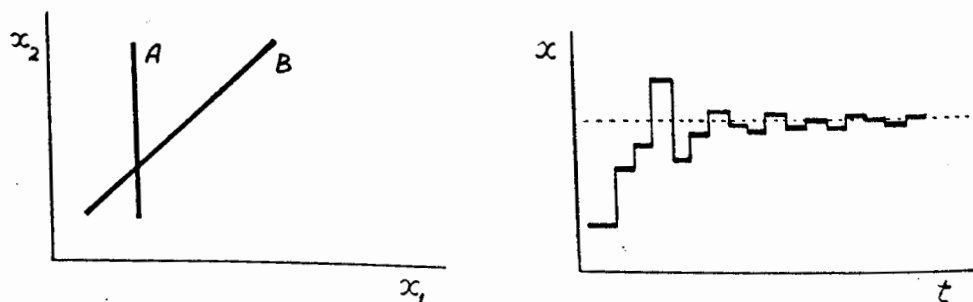


Рис. 1. Иллюстрации к методу активного поиска фонемной границы. Объяснения в тексте.

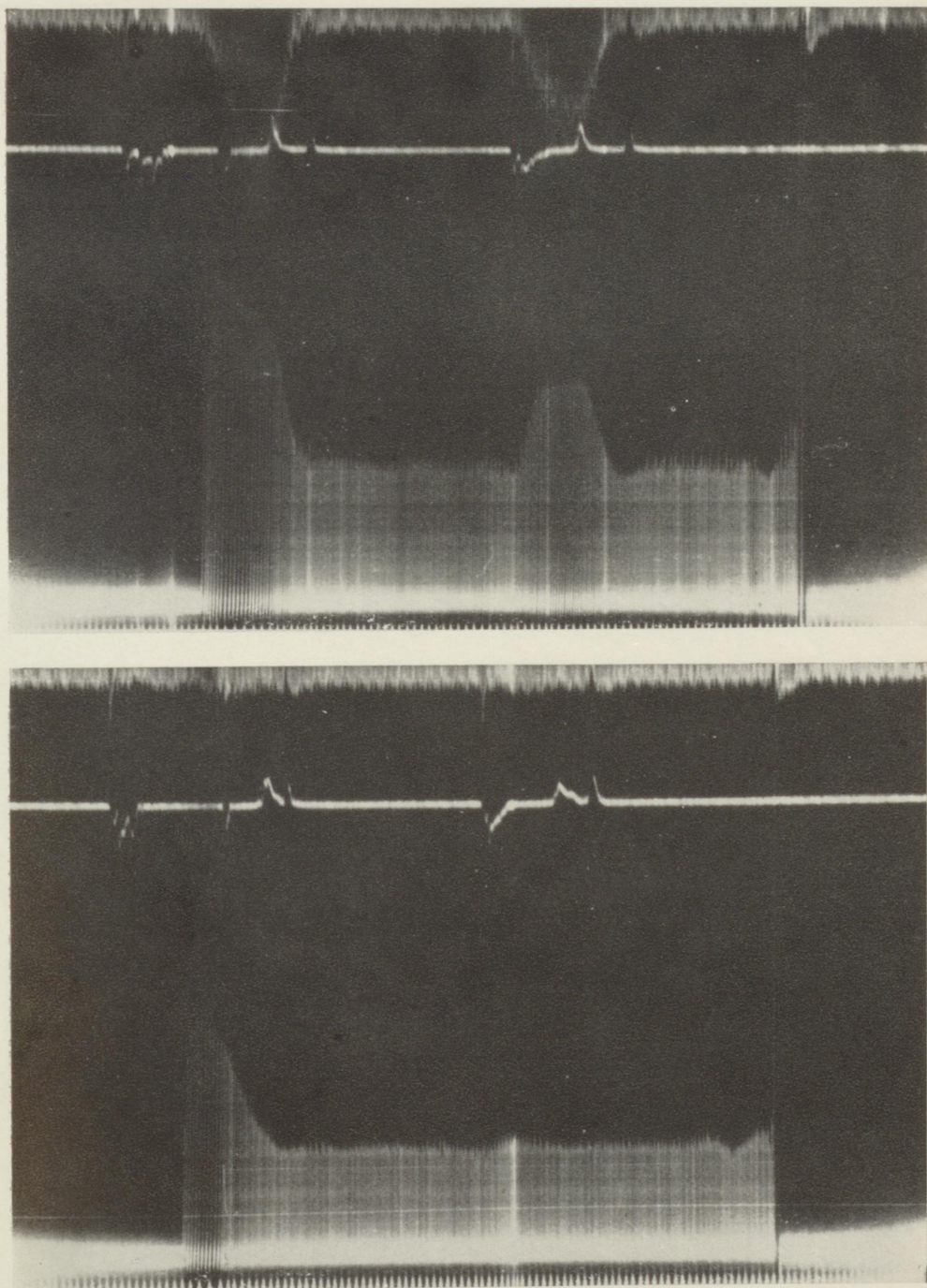


Рис. 2. Изменения периода основного тона (А) и внутриротового давления (Б) на протяжении слов *баба* и *мама*.

метру фиксировано. Другая траектория соответствует случаю, когда одновременно управляются оба параметра, связанные в данном примере линейной зависимостью.

На рис. 1 вправо показано, как во времени происходит процесс поиска границы. Его особенностью является то, что пока стимул находится в пределах одной фонемной категории, испытуемый изменяет значения параметра крупными шагами. Как только стимул пересекает фонемную границу, шаг изменения параметра резко сокращается. Испытуемый многократно проходит границу в ту и другую сторону, стремясь определить ее положение как можно точнее. Опыт показывает, что для одного определения границы испытуемому требуется прослушать 30—50 стимулов, на что уходит не более одной — двух минут.

После того, как испытуемый нашел границу и прекратил поиск, экспериментатор записывает значения параметров стимула. Для этого необходимо, чтобы в синтезаторе имелись достаточно точные индикаторы, с которых можно было бы прямо считывать значения параметров стимула.

Определение границы следует повторять несколько раз. В качестве окончательного результата эксперимента можно использовать среднее значение положения границы и среднее квадратическое отклонение отдельных определений. Последний показатель также представляет значительный интерес, т. к. он несомненно отражает точность слухового измерения параметра и стабильность критерия, применяемого при принятии решения.

Дальше я приведу некоторые данные, полученные методом поиска границ. Первая попытка применить этот метод была сделана в лаборатории Фанта — с помощью синтезатора OVE-1b мы пытались определить границы между гласными в двухформантной плоскости (2). Дальнейшие эксперименты были выполнены в Ленинграде с помощью установки, позволяющей получать очень маленький набор речеподобных стимулов. Синтезатор СГ слогов с ручным управлением находится пока в стадии изготовления. Основной целью экспериментов было накопление опыта по работе с данным методом.

Исследовалась граница между фонемами *б* и *м* в синтетических слогах *ба* и *ма*. Слог образовывался как последовательность из двух непосредственно примыкающих друг к другу „стационарных“ звуков. Источниками сигналов служили два генератора сложных гармонических напряжений, подключенные на вход двухканального электронного ключа. Преимуществом такого синтеза, при всей его примитивности, является возможность точного задания основной частоты, спектра, длительности и суммарной интенсивности для каждого из двух сигналов в последовательности.

Известно, что естественные *ба* и *ма* отличаются друг от друга по ряду акустических признаков. Одним из таких признаков, который, как ранее предполагалось, не используется человеком при восприятии, является

динамика изменения основного тона. На рис. 2 показано изменение периода основного тона из протяжении монотонно произнесенных слов *баба* и *мама*. Можно видеть, что на интервокальном *б* период основного тона увеличивается вместе с повышением внутриротового давления (верхняя кривая на нижнем снимке); при переходе к гласному он скачком возвращается к прежнему уровню. На интервокальном *м* частота основного тона или не меняется или изменяется очень незначительно.

В опытах с синтетическими слогами было сначала проверено предположение, что скачек основного тона при переходе от согласного к гласному используется для различения *б* и *м*. Исходные стимулы подбирались так, чтобы при одинаковой основной частоте на участках гласного и согласного они максимально напоминали слог *ма*. Затем стимулы изменялись так, что или понижалась основная частота на согласном или повышалась основная

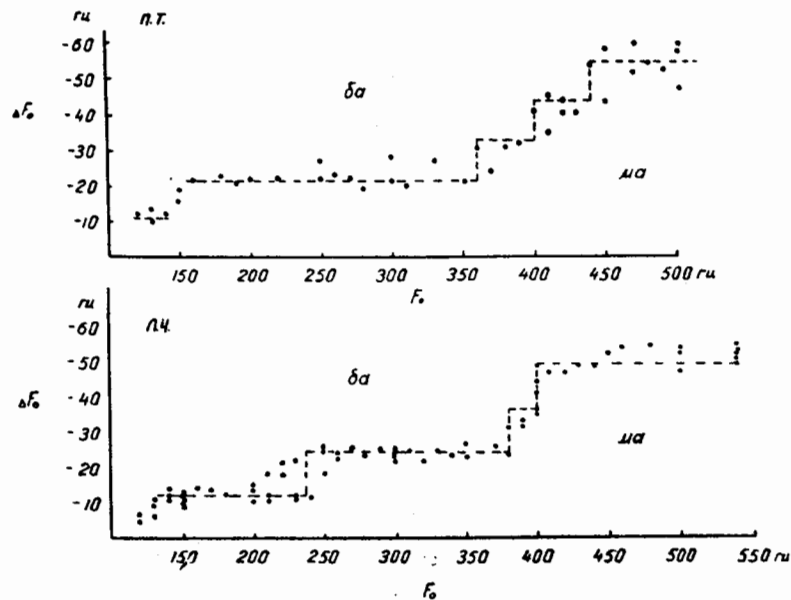


Рис. 3. Граница между *ба* и *ма* по признаку изменения основной частоты. Данные 2х испытуемых.

частота на гласном. Оказалось, что при этом стимулы начинали восприниматься как *ба*. Далее подробно исследовалось на четырех испытуемых какое значение изменения частоты при переходе от согласного к гласному соответствует границе между *м* и *б*. Измерения производились в широком диапазоне основных частот (3). Данные двух испытуемых приведены на рис. 3. По оси абсцисс отложены значения основной частоты на участке гласного в синтетическом слоге. По оси ординат — величина прироста основной частоты при переходе от согласного к гласному. В этих опытах

основная частота на гласном задавалась экспериментатором. Испытуемый изменял значение основной частоты на участке согласного, ища границу между *м* и *б*. Каждая точка на графике представляет среднее арифметическое из 5—10 определений границы. Видно, что величина изменения основной частоты, соответствующая границе, растет с увеличением абсолютного значения основной частоты, однако зависимость не является монотонной, а скорее имеет ступенчатый характер. Среднее квадратическое отклонение отдельных определений границы весьма мало на интервале, соответствующем середине ступеньки (3—4 гц), на краях оно резко возрастает.

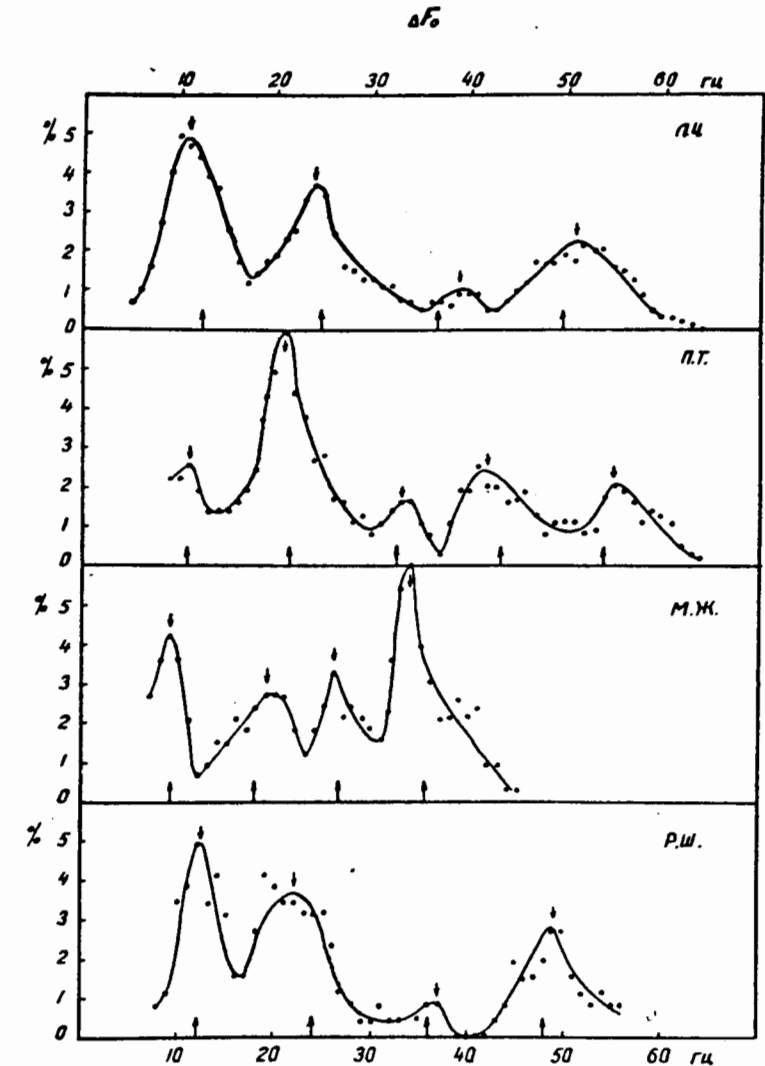


Рис. 4. Суммарные распределения определений границы между *ба* и *ма*. Данные 4х испытуемых.

Интересно, что значение ΔF_0 на второй ступеньке оказалось вдвое большим значения ΔF_0 на первой ступеньке. Это заставило обратиться к суммарным распределениям всех отдельных определений границы, полученных в опытах. Они приведены на рис. 4. Видно, что распределения имеют несколько пиков, расположенных примерно на равных расстояниях друг от друга. Причем это расстояние совпадает со значением ΔF_0 для первого пика.

Полученные данные можно объяснить с помощью представления о временном механизме измерения основной частоты звука слуховой системой. Предположим, что измеряемой мозгом величиной является интервал времени между соседними нервными импульсами, который, в зависимости от частоты сигнала равен $1, 2 \dots, k$ периодам основного тона. (Известно, что благодаря рефрактерности первый слуховой нейрон работает как делитель частоты.) Тогда для всего диапазона основных частот достаточно допустить одно значение порога, разделяющего *м* и *б*, однако это значение относится уже к частоте импульсации, а не прямо к частоте сигнала.

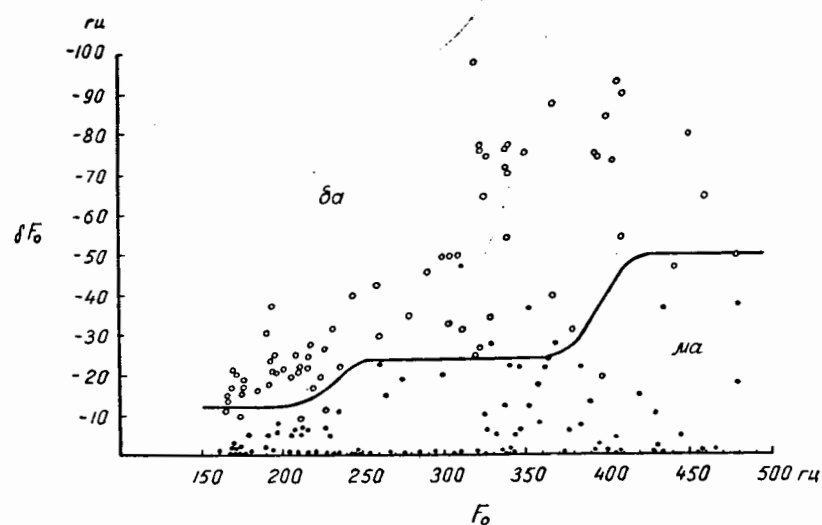


Рис. 5. Распределение естественных *ба* и *ма* в пространстве.

Очевидно, что исследование восприятия и исследование характеристик естественных звуков речи должны дополнять друг друга. После того, как в опытах с синтезом найдена граница между фонемами, которую использует человек при восприятии, целесообразно проверить, насколько хорошо эта граница разделяет естественные звуки речи.

Для проверки эффективности границ по изменению основной частоты, был использован следующий материал. Шесть дикторов наговорили монотонно на разных основных частотах слова *мама* и *баба*. Измерялись основная частота на втором гласном и изменение основной частоты при переходе

от интервокального согласного к гласному. На рис. 5 точками показаны данные касающиеся *м*, кружками — *б*. Кривая соответствует границе, определенной у одного из испытуемых. Можно видеть, что эта граница достаточно хорошо разделяет естественные *м* и *б* (правильное разделение в 94 % случаев). Худший из четырех полученных вариантов границы обеспечивал правильное разделение только в 83 % случаев.

Отсюда следует довольно тривиальный вывод, что изменение основной частоты не может быть единственным полезным признаком, различающим *б* и *м*.

Дальше я хочу рассказать о попытке использовать метод поиска границ для выяснения вопроса о том, как при восприятии объединяется информация, заключенная в различных акустических признаках сигнала, но имеющая один и тот же смысл.

Известно, что любые две фонемы различаются друг от друга более, чем по одному акустическому признаку. Казалось бы рациональным, чтобы граница между фонемами была в этом случае функцией от всех этих полезных акустических параметров. Иначе говоря, граница между фонемами А и В по параметру x_1 , должна, казалось бы, зависеть от того, какие значения имеет сигнал по параметрам x_2, x_3 . Однако, хотя это и разумно с точки зрения статистической теории распознавания образов, отнюдь не ясно способна ли нервная система реализовать сложные решающие функции. Можно предположить более простое решение вопроса, когда по каждому из акустических параметров принимаются первичные независимые решения, дальше сумма этих решений используется для выбора фонемы.

Были приведены эксперименты (4), в которых менялись два параметра, различающие *б* и *м*. Одним из них было изменение основной частоты при переходе от согласного к гласному, другим — спектральная характеристика согласного. Спектр согласного был упрощенным: он состоял из первой гармоники, равной 250 гц и четвертой гармоники, равной 1000 гц. Уровень интенсивности четвертой гармоники был переменным. В одной серии измерений экспериментатор устанавливал различные значения I_4 уровня четвертой гармоники в спектре согласного, испытуемый изменял значения основной частоты гласного, ища границу между *б* и *м*. В другой серии экспериментатор задавал значения основной частоты гласного, испытуемый менял уровень четвертой гармоники, ища ту же границу.

Полученные результаты приведены на рис. 6. По оси абсцисс отложен уровень интенсивности четвертой гармоники в спектре согласного, по оси ординат — величина изменения основной частоты при переходе от согласного к гласному. Ожидалось, что граница между фонемами будет функцией обоих параметров и будет представлена некоторой кривой, расположенной под углом к координатным осям. Как можно видеть, это предположение не оправдалось. Вместо ожидаемой одной границы, разделяющей плоскость

на две фонемные области, оказались две независимых границы, каждая по своему параметру; и пространство разделилось на четыре области.

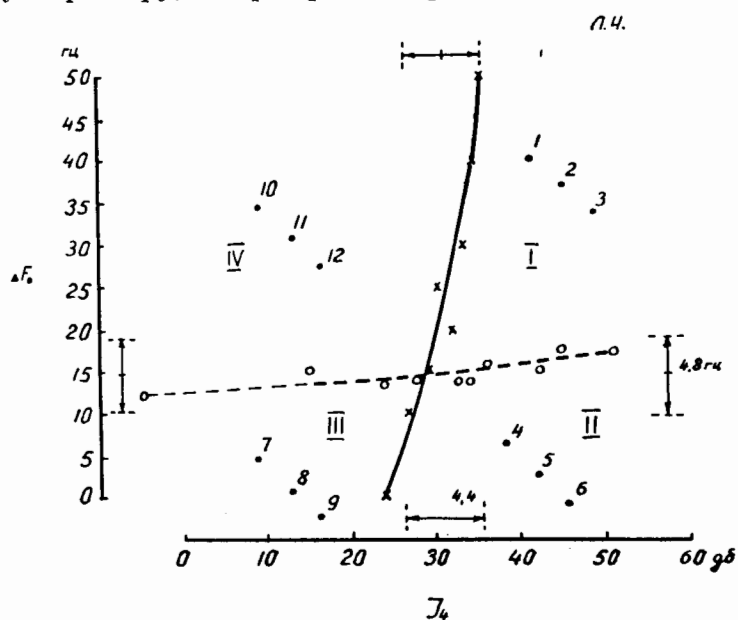


Рис. 6.

Картина оказалась аналогичной у всех трех испытуемых, только положение границ было несколько разным. На рис. 7 приведены данные другого испытуемого.

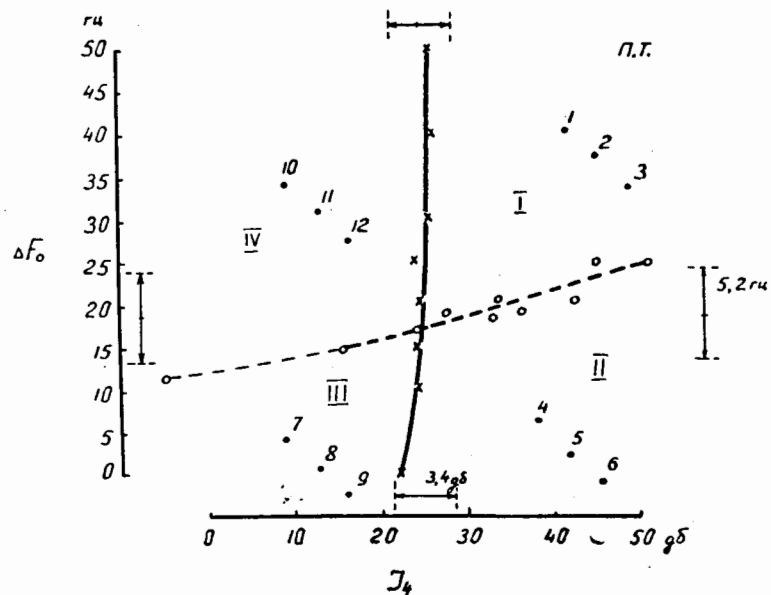


Рис. 7.

Полученные данные свидетельствуют, что по каждому параметру есть свой порог, и решение о том, по какую сторону от порога находится сигнал по одному параметру, принимается независимо от того, каковы значения сигнала по второму параметру. Иначе говоря, мы как будто бы имеем дело с системой, состоящей из параллельных каналов с фиксированными решающими правилами в каждом из каналов. Однако, оказывается также, что окончательная фонемная интерпретация стимулов не является постоянной. Стимулы из областей 1 и 3 в одной серии измерений воспринимались как б, а в другой — как м. Когда испытуемый менял уровень интенсивности четвертой гармоники и сравнивал между собой сигналы областей 4 и 1,

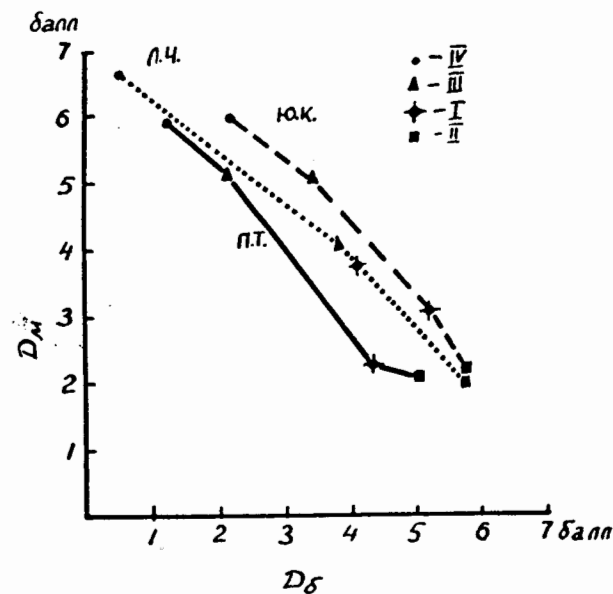


Рис. 8.

он трактовал стимулы области 1 как м. Когда он менял величину скачка основной частоты и сравнивал сигналы областей 2 и 1, он трактовал стимулы области 1 как б. Если мы имеем дело со схемой, принимающей параллельные независимые решения по отдельным акустическим параметрам, возможность различной интерпретации стимулов областей 1 и 3 представляется довольно логичной. Это действительно сигналы, которые по одному признаку являются фонемой б, а по другому — фонемой м.

Здесь однако возникает весьма интересный вопрос. Являются ли различные признаки равноценными, или один имеет больший вес, чем другие при окончательной фонемной интерпретации стимула? Ответить на этот вопрос методом исследования фонемных границ уже нельзя. Была сделана

попытка применить для этой цели метод шкалирования субъективных расстояний между стимулом и фонемой. Значения выбранных 12 стимулов показаны точками на рис. 7. Стимулы были записаны на пленку в случайной последовательности. Они предъявлялись тем же трем испытуемым. В одной серии прослушиваний испытуемые определяли для каждого стимула его расстояние в баллах от фонемы *б*, в другой серии — от фонемы *м*.

Полученные результаты приведены на рис. 8. Даны средние расстояния в баллах для стимулов каждой из четырех областей. По оси абсцисс отложено расстояние от фонемы *б*, по оси ординат — от *м*. Можно видеть, что стимулы области 4 оцениваются как близкие к *б* и далекие от *м*. Стимулы области 2 близки к *м* и далеки от *б*. Стимулы областей 1 и 3 занимают промежуточное положение, однако стимулы области 1 ближе к *м*, а стимулы области 3 ближе к *б*. Это уже говорит о том, что вес спектрального признака согласного больше веса признака скачка основной частоты.

Для количественного вычисления весов было сделано предположение, что субъективное расстояние между стимулом и фонемой пропорционально разности между некоторым постоянным числом, представляющим идеальную фонему, и суммой весов положительных признаков данной фонемы, присутствующих в стимуле. Оказалось, что относительные веса признаков довольно сильно различались у наших трех испытуемых. Так, у одного испытуемого вес признака скачка основной частоты составил примерно 0,80 от веса спектрального признака, тогда как у другого испытуемого он составил всего 0,16 от веса спектрального признака.

Настоящие данные слишком ограничены, чтобы на их основании можно было делать какие либо обобщения относительно истинной организации процесса восприятия. Однако, я все же рискну высказать некоторую гипотезу. Она состоит в том, что процесс принятия фонемных решений состоит из двух этапов. На первом этапе принимаются решения об отдельных полезных акустических признаках, устанавливается, в какой области значений по каждому из параметров находится сигнал. На этом этапе распознающая система действует как система с постоянными характеристиками: и способ измерения параметра и решающие правила фиксированы. На следующем этапе происходит суммирование первичных решений, помноженных на соответствующие веса, и сравнение их с некоторыми порогами. Логично предположить, что значения весов устанавливаются в процессе обучения, а значения порогов регулируются текущим образом в зависимости от контекста и других априорных сведений о том, какая фонема должна быть на данном месте. Такая гипотеза, напрашивающаяся на основании полученных данных, представляется достаточно физиологичной, т. к. соответствующая ей модель может быть довольно просто реализована с помощью нервной сети. Если эта модель соответствует истине, метод поиска границ годится для исследования первого этапа распознавания: вы-

явления и уточнения природы измеряемых слухом признаков сигнала и применяемых к ним решающих правил. Для исследования второго этапа необходимы другие методы.

ЛИТЕРАТУРА

1. Fant, G.: The nature of distinctive features. *STL-QPSR* 4/1966, 1—14.
2. Chistovich, L.—Fant, G.—de Serpa-Leitao, A.: Mimicking and perception of synthetic vowels. Part II. *STL-QPSR* 3/1966, 1—3.
3. Чистович Л. А.: Изменение основной частоты голоса как различительный признак согласных. *Акустический журн.* В печати.
4. Чистович Л. А.: О процедуре распознавания фонем человеком. *Вопросы психологии.* В печати.

DISCUSSION

Fischer-Jørgensen:

I find Dr. Chistovich's experiments very interesting. I only want to emphasize (and I do not think I really disagree with her on this point) that what is found by this method is not a *phoneme* boundary, but a boundary between two areas in the phonetic space which in this particular environment and in this particular situation are conceived as manifestations of two different phonemes. As the boundaries are very often different for different environments, we shall not expect to be able to find clear boundaries between areas covering all manifestations of given phonemes or distinctive features, but rather cases of complex overlappings. This was shown thirty years ago for German vowel quantity by E. Zwirner.

Lindner:

Die vorgetragenen Untersuchungen über die Bestimmung der Grenzen zwischen Klassen von Lauten decken sich im wesentlichen mit Untersuchungen, die in den letzten beiden Jahren an unserem Institut durchgeführt worden sind. Dabei zeigte es sich, daß die Grenzen zwischen Vokalklassen von der Herkunft des Perzipienten und seiner Zugehörigkeit zu einer Dialektgemeinschaft abhängig sind. Neben objektiv bestimmbareren Parametern des akustischen Signals müssen auch subjektiv bedingte Parameter berücksichtigt werden.

Tillmann:

I wonder whether decision theory by its purely statistical nature could help us much in finding a model of speech perception, even a neurological one. Decision theoretic work on pattern recognition has also shown empirically what is logically to be proved: that decision theory is blind with regard to certain kinds of structure.

Fujisaki:

I should like to express my respect for the extensive investigation which Dr. Chistovich has just reported, particularly because I myself have been conducting research on the perceptual boundaries between Japanese vowels. May I make two comments:

(I) I believe that the experiments on the boundary between (ba) and (ma) would have been more meaningful if the two parameters were varied not independently, but in accordance with relationships found in natural speech.

(II) I should like to ask if Dr. Chistovich would present a physiological model which would support the particular shape of the boundary obtained in the F_0 — ΔF_0 plane.

Загоруйко:

Я хотел бы выступить в защиту статистической теории распознавания образов. В терминах этой теории можно объяснить все те интересные экспериментальные данные, которые получены в работе Л. А. Чистович. В частности, можно легко показать, что механизм принятия решения на основании порогового измерения отдельных признаков проще (экономичнее), чем с помощью многомерной разделяющей гиперповерхности.