

173

АКАДЕМИЯ НАУК СССР

АНАЛИЗ
РЕЧЕВЫХ
СИГНАЛОВ
ЧЕЛОВЕКОМ

АКАДЕМИЯ НАУК СССР
НАУЧНЫЙ СОВЕТ
ПО ФИЗИОЛОГИИ ОРГАНОВ ЧУВСТВ

АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ ЧЕЛОВЕКОМ



ИЗДАТЕЛЬСТВО «НАУКА»
Ленинградское отделение
Ленинград · 1971

ОТВЕТСТВЕННЫЙ РЕДАКТОР
чл.-корр. АН СССР *Г. В. ГЕРШУНИ*

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

*А. В. Бару, А. В. Венцов, В. И. Галунов,
С. Н. Гольдбург, В. А. Кожевников, Ю. И. Кузьмин,
А. П. Молчанов, Е. А. Радионова,
Л. А. Чистович (зам. отв. редактора)*

РЕДАКТОРЫ ТОМА:

*А. В. Венцов, В. А. Кожевников,
Ю. И. Кузьмин, Л. А. Чистович*

ПРЕДИСЛОВИЕ

Настоящее издание является VII томом сборников «Проблемы физиологической акустики», которые начали выходить с 1949 г. под общей редакцией академика Л. А. Орбели. В данном томе — «Анализ речевых сигналов человеком» — собраны работы, продолжающие направление исследований, отраженное в V томе «Проблем физиологической акустики», носящем название «Механизмы речеобразования и восприятия сложных звуков», и в книге «Речь. Артикуляция и восприятие» (Л., 1965).

Публикуемые статьи связываются общей системой взглядов относительно структуры процесса восприятия речи и относительно путей подхода к экспериментальной разработке проблемы. В более полном виде эта система представлений изложена в коллективной работе Л. В. Бондарко, Н. Г. Загоруйко, В. А. Коженикова, А. П. Молчанова, Л. А. Чистович «Модель восприятия речи человеком» (Новосибирск, 1968).

Основное предположение заключается в том, что процесс восприятия имеет многоуровневую структуру. На начальном этапе обработки — этапе «слухового анализа» — производится обнаружение и измерение определенных параметров, признаков сигнала. Для обозначения характеристик сигнала по этим признакам в статьях сборника применяется термин «слуховое описание сигнала». Разрабатываемая модель восприятия речи предполагает, что на следующих этапах преобразования информации осуществляется переход к такой сокращенной и абстрактной форме описания, которая допускает запоминание и дальнейший лингвистический анализ достаточно длинных участков речевого сообщения, и может непосредственно использоваться для управления процессами речеобразования при воспроизведении сигнала. Этот уровень описания условно назван фонетическим описанием, а процесс перехода от слухового к фонетическому описанию определен как фонетическая интерпретация.

Все статьи сборника группируются вокруг вытекающих из вышесказанного двух основных проблем. Одна из них касается слухового описания и состоит в определении признаков сигнала, измеряемых слухом, и в разработке моделей, обеспечивающих осуществление необходимых измерений. Вторая проблема касается структуры фонетического описания и процедуры фонетической интерпретации.

ПОЯСНЕНИЯ К ТРАНСКРИПЦИИ

Во всех работах сборника для обозначения звуков использовались знаки международной фонетической транскрипции. Следует, однако, отметить, что при обозначении синтетических звуков не преследовалась цель отразить их звучание наиболее точно. В этих случаях символика употреблялась для обозначения не фонетического качества отдельного звука, а для обозначения способа интерпретации слушателями целого континуума сигналов. Таким образом, транскрипция была не столько фонетической, сколько фонематической.

Особых пояснений требуют следующие обозначения. Знак [t̚] был принят для обозначения русского заднеязычного гласного ы; для обозначения мягкости согласного использовался штрих, расположенный сверху справа от основного символа, например [lʰ].

При цитировании работ других авторов сохранялась символика оригинала.

В тех случаях, когда слушатели записывали результаты своей оценки сигналов русскими буквами, при описании полученных данных также использованы знаки русского алфавита.

1. СЛУХОВОЕ ОПИСАНИЕ РЕЧЕПОДОБНЫХ СИГНАЛОВ (ПОЛЕЗНЫЕ ПСИХОАКУСТИЧЕСКИЕ ПРИЗНАКИ)

О СЛУХОВОМ ОПИСАНИИ ГЛАСНОГО

1. ПРИЗНАКИ, РАЗЛИЧАЮЩИЕ [i] и [e]

В. Н. Мушников, Л. А. Чистович

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

Одной из важнейших проблем речевой психоакустики является определение признаков, используемых слуховой системой для описания сигнала со сложным спектром. Существующие в настоящее время гипотезы, нашедшие свое отражение в предлагаемых устройствах для автоматического распознавания речи, можно грубо разделить на две группы. Одно предположение состоит в том, что слуховое описание, например, стационарного гласного представляет собой некоторый эквивалент его энергетического спектра. Оно может быть представлено r -мерным вектором, где r — число спектральных отсчетов, определяемое количеством фильтров в гребенке фильтров, осуществляющих частотный анализ. Предполагается, что так же описаны и эталоны гласных, с которыми сравнивается входной сигнал при его распознавании.

Другая группа гипотез предполагает, что существующий на каком-то уровне обработки эквивалент спектра сигнала подвергается дополнительным преобразованиям. В результате этих преобразований выделяются определенные признаки спектра (положение максимумов на шкале частот, отношения энергий в определенных частотных полосах и т. д.), которые и используются в качестве слухового описания.

В случае правильности второй гипотезы можно предположить, что для различения определенных двух фонем, объективно отличающихся друг от друга всей формой спектра в целом, человек использует только какие-то локальные признаки спектра. Например, он различает одну пару фонем по признакам низкочастотной области спектра, а другую пару фонем — по признакам высокочастотной области. Использование локальных признаков спектра может обеспечить устойчивость восприятия к таким искажениям спектра, как плавный подъем или падение вдоль шкалы частот.

В настоящей работе в качестве примера было выбрано различие гласных [i] и [e]. Объективно эти гласные различаются друг от друга по частоте F_1 , частоте F_2 , а также по разности уровней между первой и второй формантой (Peterson, Barney, 1952; Варшавский, Чистович, 1959; Fant, 1959).

В работе применялись двухформантные синтетические гласные. Задача состояла в том, чтобы определить форму фонемной границы между [i] и [e] в F_1F_2 -плоскости и выяснить, зависит ли эта граница от разности уровней между первой и второй формантами.

МЕТОДИКА

Работа проводилась с параллельным формантным синтезатором. Постоянными параметрами стимула были $F_0=114$ гц (в эксперименте осуществлялся постоянный контроль за основной частотой, ее отклонение не превышало ± 1 гц), форма спектра источника голоса (наклон огибающей спектра составлял -6 дб на октаву в сторону высоких частот), длительность, равная 300 мсек.

Переменными параметрами стимула, управляемыми экспериментатором, были или частота F_2 , или уровень интенсивности второй форманты. Исследование влияния уровня проводилось при одном фиксированном значении $F_2=2250$ гц.

Испытуемый управлял значением F_1 , его задача состояла в том, чтобы найти границу между [i] и [e].

Изменение F_1 осуществлялось путем управления емкостью в цепи формантного фильтра. Испытуемый управлял одним из параллельно включенных магазинов емкостей. Значение на другом магазине, неизвестное испытуемому, устанавливалось экспериментатором. Минимальный шаг по частоте F_1 в области около 400 гц, где оказалась расположенной граница, составлял 1.5 гц.

В опытах участвовало 3 испытуемых. Каждый из них производил по 10 повторных определений границы для каждого значения переменного параметра.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На рис. 1 и 2 точками показаны средние значения границы по F_1 по объединенным данным всех трех испытуемых. Можно видеть, что экспериментальные зависимости проще всего аппроксимируются прямыми, т. е. оказывается, что граница между [i] и [e] определяется только F_1 и не зависит ни от частоты F_2 , ни от разности уровней между первой и второй формантами.

Независимость границы от F_2 подтверждает те предварительные результаты, которые были получены в лаборатории Фанта (Chistovich et al., 1966) при исследовании границ между гласными в плоскости F_1F_2 . Из 102 найденных тогда границ 80 аппроксимировались прямыми, параллельными оси F_2 или F_1 . К сожалению, в этих измерениях были возможны значительные систематические ошибки, не позволявшие сделать выводы об абсолютных значениях частот формант, соответствующих границам. Полезным методическим результатом этих предварительных экспериментов было то, что положение границы не зависит от траектории в про-

пространстве F_1F_2 , по которой испытуемый смещает сигнал в процессе поиска границы. Это позволило в настоящей работе ограничиться случаем, когда значение F_2 на протяжении поиска границы является фиксированным.

За последние годы в нескольких работах были получены точки в пространстве F_1F_2 , соответствующие границам между [i] и [e] или [i] и [e]. В этих работах (Janota, 1967; Stevens et al., 1969; Fujisaki, Kawashima, 1968; Голузина, 1971) исследовалась фонемная идентификация гласных группой испытуемых; траектории в формантном пространстве, по которым были расположены применявшиеся в опытах стимулы, в разных работах были различными.

На рис. 3 эти траектории нанесены прерывистыми линиями. Точками

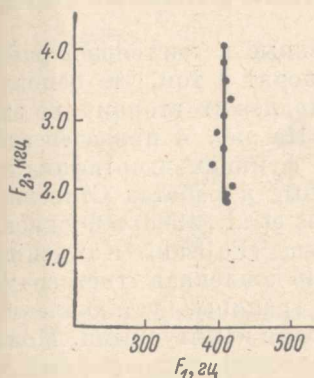


Рис. 1. Среднее значение границы между [i] и [e] по частоте первой форманты (F_1) при различных значениях частоты второй форманты (F_2).

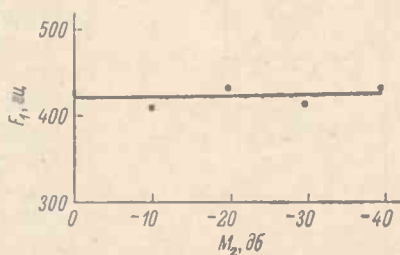


Рис. 2. Зависимость границы по F_1 от уровня интенсивности второй форманты при $F_2=2250$ гц.

По оси абсцисс — затухание в канале F_1 ;
по оси ординат — частота F_1 .

показаны значения найденных границ. Сплошная прямая указывает положение границы, полученное в нашей работе. Можно видеть, что данные очень близки по значениям F_1 (несколько отличаются данные Янота). Это подтверждает вывод о том, что различие гласных [i], [i] и [e] основывается на признаках спектра в низкочастотной области.

Интересен тот факт, что положение границы оказывается одним и тем же у носителей различных языков. Исследование Стивенса и других проводилось на группе американцев и группе шведов. Положение границы у этих групп совпало, оно представлено одной точкой на рис. 3. В работе Голузиной, как и в нашей работе, испытуемыми были русские, в работе Янота — чехи, а в работе Фужизаки и Кавашима — японцы.

Вероятное предположение состоит в том, что положение границы отражает какую-то особенность организации слухового описания. В работе Стивенса и других было обнаружено, что раз-

личимость гласных стимулов, определенная АВХ-методом, имеет отчетливый максимум вблизи данной границы.

В экспериментах по идентификации гласных обычно выявляется довольно широкая переходная область — интервал частот, внутри которого вероятность опознания одной из фонем растет от 0 до 1, а вероятность опознания другой фонемы соответственно падает.

Наличие такой достаточно широкой области может объясняться как несовпадением положения фонемных границ у разных испытуемых, так и флюктуациями границы у одного и того же испытуемого.

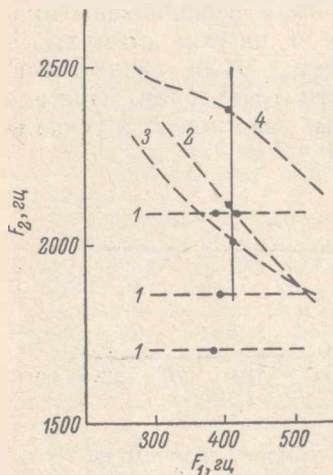


Рис. 3. Положение границы между [i] и [e], полученное методом фонемной идентификации, по данным разных авторов.

1 — Janota, 1967; 2 — Fujisaki, Kawashima, 1968; 3 — Stevens et al., 1969; 4 — Голузина, 1971.

Полученные в настоящей работе данные говорят о том, что основная роль принадлежит второму из этих факторов. На рис. 4 приведены для сравнения функции идентификации, определенные в работах Стивенса и др. (Stevens et al., 1969) и Фужизаки и Кавашима (Fujisaki, Kawashima, 1968), и накопленная гистограмма значений границы, установленных нашими тремя испытуемыми. Можно видеть, что совпадение данных более чем удовлетворительно. На этом же рисунке показаны для трех наших испытуемых интервалы частот, в которых располагались 80% из всех установленных ими границ. Можно видеть, что эти интервалы сильно перекрываются.

Очень интересным оказалось то, что граница может сохранять устойчивое положение в течение довольно длительного времени. Об этом

говорит тот факт, что значения стандартного отклонения для серии из 10 последовательных установок границ оказались в ряде случаев весьма малыми. Так, в 17 случаях из 46 они не превышали 12 Гц, т. е. 3% от среднего значения. По данным Фланагана (Flanagan, 1955), дифференциальный порог по частоте форманты, определяемый в условиях АВХ-сравнения, составляет 3—5%, т. е. имеет примерно ту же величину.

Полученные данные не позволяют ответить на вопрос, с чем связаны флюктуации границы, однако имеется некоторый намек на то, что граница менее устойчива в тех случаях, когда гласный теряет естественность (большие значения частоты F_2 , слишком низкий или слишком высокий уровень второй форманты). На двух

из 6 полученных графиков зависимости величины стандартного отклонения границы от частоты и уровня второй форманты эта тенденция была выраженной. Эти два случая приведены на рис. 5.

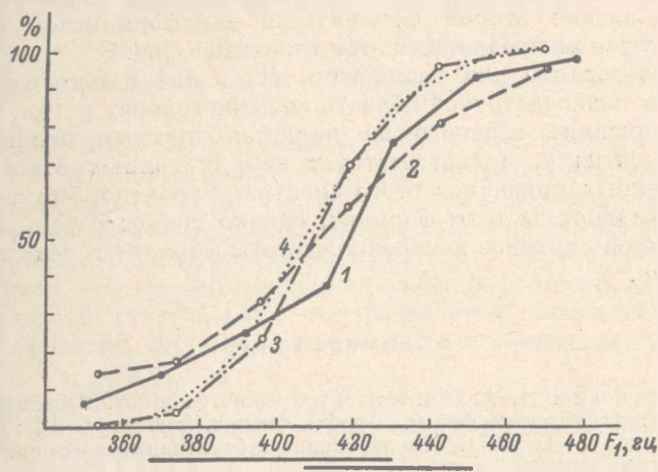


Рис. 4. Кривые фонемной идентификации, полученные в работах разных авторов.

1 — для японцев (Fujisaki, Kawashima, 1968); 2 и 3 — для американцев и шведов соответственно (Stevens et al., 1969); 4 — гистограмма, полученная в настоящей работе для трех русских испытуемых (линии под осью абсцисс показывают для каждого испытуемого области, в пределах которых лежат 80% определений границы).

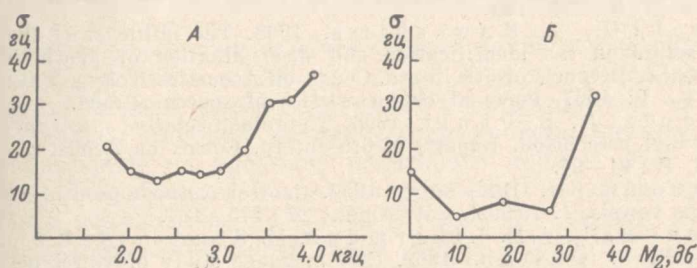


Рис. 5. Значения стандартного отклонения границы по F_1 при разной частоте (А, испытуемый В. М.) и при разном уровне интенсивности второй форманты (Б, испытуемый Л. Ч.).

ЗАКЛЮЧЕНИЕ

Полученные в работе данные показывают, что различие [i] и [e] основывается на частоте первой форманты гласного. Ни частота второй форманты, ни ее уровень не учитываются при различении этих гласных между собой. Отсюда, естественно,

не следует, что эти параметры вообще не имеют значения для распознавания этих гласных. Из ряда данных (Fujimura, 1967; Lindquist, Pauli, 1968; Stevens et al., 1968) следует, что частота F_2 используется для различения [i] от [y] или [ü]. Известно также, что подавление второй форманты в двухформантном синтетическом гласном превращает этот гласный в [u].

Использование для различения [i] и [e] только их особенностей в низкочастотной области спектра говорит о том, что слуховое описание гласного не является простым эквивалентом спектра сигнала, а соответствует набору данных относительно определенных признаков этого спектра. Вероятно, что эти признаки отражают частоты формант, однако вопрос о том, как осуществляется слуховое измерение частоты форманты, пока остается открытым.

Л и т е р а т у р а

- В а р ш а в с к и й Л. А., Л. А. Ч и с т о в и ч. 1959. Средние спектры русских гласных фонем. Пробл. физиол. акустики, IV : 181—186.
- Г о л у з и н а А. Г. 1971. Шкалирование субъективных расстояний между синтетическими гласными. Наст. сб. : 155—160.
- С h i s t o v i c h L., G. F a n t, A. d e S e r p a - L e i t a o. 1966. Miming and perception of vowels. STL—QPSR, 3: 1—3.
- F a n t G. 1959. Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics, 15, 1 : 1—106.
- F l a n a g a n J. L. 1955. A difference limen for vowel formant frequency. J. Acoust. Soc. Amer., 27 : 613—617.
- F u j i m u r a O. 1967. On the second spectral peak of front vowels: a perceptual study of the role of the second and third formants. Language a. Speech, 10 pt. 3 : 181—193.
- F u j i s a k i H., T. K a w a s h i m a. 1968. The influence of the various factors on the identification and discrimination of synthetic speech sounds. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 95—98.
- J a n o t a P. 1967. Personal characteristics of speech. Praha.
- L i n d q u i s t J., S. P a u l i. 1968. The role of relative spectrum levels in vowel perception. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 91—94.
- P e t e r s o n G., H. B a r n e y. 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Amer., 24 : 175—184.
- S t e v e n s K. N., A. M. L i b e r m a n n, M. S t u d d e r t - K e n n e d y, S. E. G. Ö h m a n. 1969. Crosslanguage study of vowel perception. Language a. Speech, 12 : 1—23.
-

О СЛУХОВОМ ОПИСАНИИ ГЛАСНОГО

II. ОБНАРУЖЕНИЕ ВТОРОЙ ФОРМАНТЫ В СИНТЕТИЧЕСКОМ ГЛАСНОМ

В. Н. Мушников, Л. А. Чистович

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

Распространено предположение, что слуховое описание стационарного гласного соответствует указанию положения по оси частот (проекции улитки) некоторого числа максимумов на отклике улитки на этот гласный. Сколько должно быть таких максимумов, чему в спектре гласного они должны соответствовать (гармоникам, формантам, группам из соседних формант), насколько они должны быть выражены — определяется параметрами ряда звеньев слуховой системы, которые пока изучены еще слишком мало.

Один из возможных психоакустических подходов к решению вопроса состоит в том, чтобы определить, какими особенностями должен характеризоваться спектр синтетического гласного, для того чтобы человек мог обнаружить присутствие в этом гласном определенной форманты.

При выборе методики исследования мы исходили из следующих фактов: 1) одноформантный гласный с частотой форманты ниже 700 гц воспринимается как [u] или [o] (Варшавский, Литвак, 1955); 2) подавление второй и более высоких формант в спектрах естественных [i], [e], [o], [ü] превращает эти гласные в [u] или [o].

Отсюда следовало ожидать, что синтетический двухформантный гласный с $F_1 < 700$ гц и 1200 гц $< F_2 < 2400$ гц с фиксированной по уровню первой формантой будет восприниматься как [u] или [o] только при том условии, если уровень интенсивности второй форманты не превысит некоторого порога. Этот порог, при котором произойдет изменение фонемного качества гласного (он начнет походить на [e], [i], [ö] или [ü]), можно принять за порог обнаружения присутствия второй форманты.

Гласный с пороговым уровнем второй форманты будем называть пороговым по A_2 . Располагая спектрами ряда пороговых по A_2 гласных, систематически отличающихся друг от друга по таким параметрам, как частота F_2 , наклон спектра в сторону высоких частот, можно, вероятно, составить некоторые представления о том, что принимается слухом за спектральный максимум.

МЕТОДИКА

Первая часть работы проводилась на параллельном формантном синтезаторе, упрощенная блок-схема которого приведена на рис. 1.

На рис. 2 приведены спектры (с учетом частотной характеристики телефона) двух применявшихся в опытах «одноформантных» стимулов (случай,

когда затухание в канале F_2 максимально). Стимулы имели одну и ту же основную частоту (300 гц), одинаковый уровень звукового давления (83 дб от $2 \cdot 10^{-4}$ дин/см²) на частоте 600 гц (частота первой форманты) и различались по характеру спада спектра в области высоких частот. Оба стимула воспринимались как [o]- или [u]-образный гласный. Частота второй форманты устанавливалась на частотах гармоник в диапазоне 1200—2400 гц. Стимулы длительностью 200 мсек. подавались на телефон. Восприятие было монауральным.

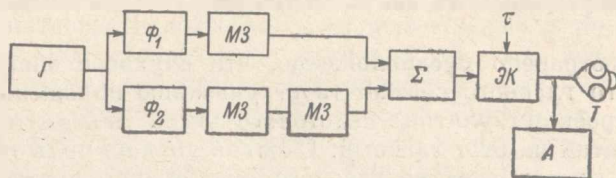


Рис. 1. Упрощенная блок-схема параллельного формантного синтезатора.

Г — генератор основного тона; Φ_1 и Φ_2 — формантные фильтры; МЗ — магазин затухания; Σ — сумматор; ЭК — электронный ключ; Т — динамический телефон ТД-6; А — анализатор звуковых частот АН-1-50; τ — длительность стимула, задаваемая схемой управления.

Задача испытуемого состояла в том, чтобы, изменяя затухание в канале F_2 , подобрать такое его пороговое значение, при котором гласный переходит из [o]—[u] в любую другую фонему. Испытуемый управлял одним из МЗ в канале F_2 , значение затухания на втором МЗ, неизвестное испытуемому, устанавливалось экспериментатором и менялось в случайном порядке от одного измерения к другому в диапазоне 10—30 дб.

В опытах участвовало двое испытуемых. Каждый из них повторял определение порога по 5 раз. За порог принималось среднее арифметическое из 10 значений, определенных обоими испытуемыми. Спектр стимула, полученного при установке этого порогового затухания в канале F_2 , измерялся анализатором АН-1-50 (полоса пропускания 15—20 гц).

РЕЗУЛЬТАТЫ

Для определения стабильности порога изменения фонемного качества звука были вычислены средние значения разности между затуханиями, устанавливаемыми при последовательных измерениях в серии, а также стандартные отклонения этих разностей. Так как средние значения разностей не отличались существенно от нуля, можно было заключить, что закономерного сдвига порога в течение серии измерений не происходит.

Значение стандартного отклонения отдельного измерения порога было вычислено путем деления стандартного отклонения разности между последовательными измерениями на $\sqrt{2}$. Полученные величины оказались близкими у обоих испытуемых (Л. Ч. — 2.50 дб, В. М. — 2.32 дб). Статистически существенных систематических различий в порогах также обнаружено не было.

Спектры полученных пороговых по A_2 гласных приведены на рис. 2. На этих спектрах действительно можно обнаружить

максимумы, соответствующие F_2 . Однако о максимуме можно говорить лишь в том смысле, что уровень составляющей с частотой F_2 превышает среднее значение уровней ближайшей левой и правой составляющих. Это превышение, как можно видеть, оказывается очень небольшим. Кроме того, разность уровней между составляющей с частотой F_2 и максимальной составляющей в спектре (600 гц) весьма значительна. Для случая $F_2=2100$ гц она составляет 46 дб. Оба эти факта говорят о том, что система, осуществляющая обнаружение максимума, должна иметь на своем входе анализатор частот с достаточно высокой разрешающей способностью.

Однако возможно, что при большой разрешающей способности анализатора система начнет обнаруживать максимумы, соответствующие каждой отдельной гармонике в спектре сигнала.

Высказывание, что гласный, пороговый по A_2 , является для восприятия двухформантным, соответствует утверждению, что в отклике слуховой системы на этот гласный выделяются два максимума. При этом подразумевается, что в отклике на «одноформантный» гласный нет максимумов, соответствующих отдельным составляющим в интервале частот между 600 и 2400 гц. Если это так, то человек не должен обнаруживать присутствие этих составляющих в спектре сигнала.

Для проверки этого предположения были проведены дополнительные эксперименты с использованием установки такого же типа, как приведенная на рис. 1, но

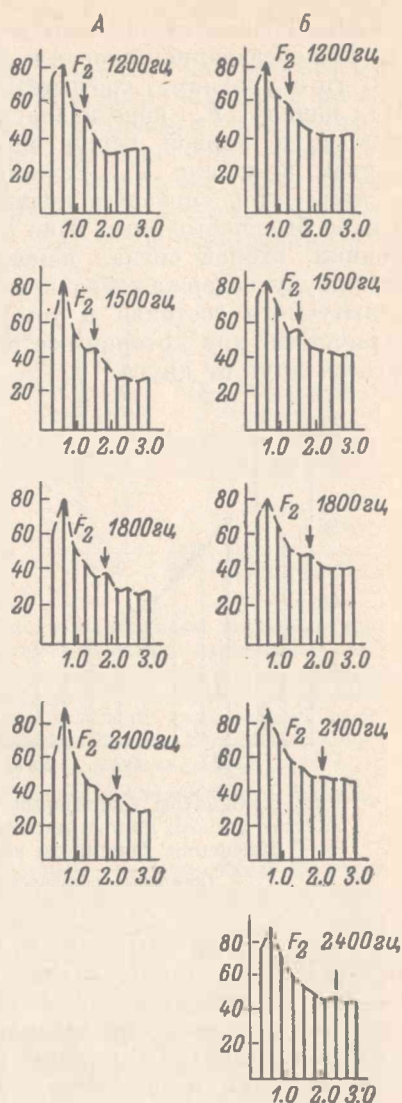


Рис. 2. Спектры синтетических гласных, пороговых по амплитуде второй форманты, при разных значениях F_2 .

По оси абсцисс — частота, кгц; по оси ординат — уровень звукового давления, дб от $2 \cdot 10^{-4}$ дин/см². А — 12 дб/окт.; Б — 6 дб/окт. Спектры А и Б отличаются крутизной спада огибающей гармоник основного тона. Стрелкой показан максимум на частоте второй форманты.

с заменой формантного синтезатора на генератор ГСН-1. По каналу F_1 передавались первая и вторая гармоники основной частоты 300 гц с уровнями, соответствующими 83 дб от $2 \cdot 10^{-4}$ дин/см². По каналу F_2 передавалась одна из верхних гармоник. Как и в первой серии опытов, испытуемый управлял затуханием в канале F_2 . Разница в постановке опыта состояла в том, что сигналы подавались парами. Первый сигнал в паре был постоянным и не содержал исследуемой верхней гармоники, второй сигнал изменялся испытуемым, регулируя амплитуду содержащейся в сигнале верхней гармоники. Задача испытуемого состояла в том, чтобы найти пороговый уровень гармоники, при котором постоянный и переменный сигналы отличаются друг от друга.

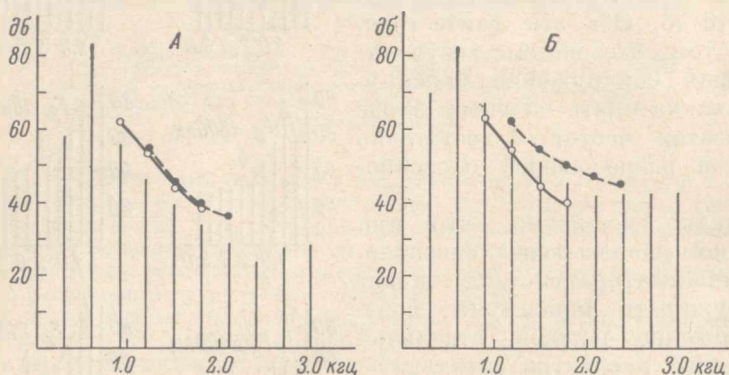


Рис. 3. Спектры «одноформантных» гласных стимулов (вертикальные линии), пороги маскировки (сплошная кривая) и пороги изменения фонемного качества (прерывистая кривая).

Остальные обозначения те же, что на рис. 2.

Были определены пороги маскировки гармоник 900, 1200, 1500 и 1800 гц. Данные приведены сплошными кривыми на рис. 3. На тех же графиках показаны значения порога изменения фонемного качества, определенные в первом эксперименте, т. е. значения амплитуд гармоник, соответствующих вторым формантным максимумам на спектрах рис. 2.

Можно видеть, что в «одноформантном» гласном со спектром А составляющие в полосе 900—1800 гц оказываются ниже порога маскировки, а порог изменения фонемного качества буквально совпадает с порогом маскировки. Следовательно, в данном случае мы действительно можем утверждать, что обнаружение второй форманты соответствует обнаружению дополнительного максимума на отклике слуховой системы на гласный стимул.

В «одноформантном» гласном со спектром Б составляющие в полосе 900—1800 гц оказались превосходящими по уровню

интенсивности порог маскировки. Это потребовало проведения дополнительных экспериментов.

Прежде всего было обнаружено, что на пороге маскировки присутствие в сигнале гармоники 900 гц воспринимается не как изменение фоновое качество, а как увеличение естественности звучания того же самого гласного. Для изменения фоновое качества необходимо дополнительное увеличение уровня гармоники. Так как порог изменения фоновое качества оказался равным 70 дб, можно было заключить, что в «одноформантном»

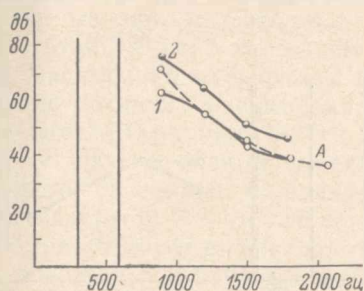


Рис. 4. Переходная область при восприятии гласного.

1 — порог маскировки гармоники; 2 — максимальный уровень гармоники, при котором гласный, состоящий из этой гармоники и двух нижних гармоник (их уровни интенсивности показаны вертикальными линиями), еще может быть принят за одноформантный. Прерывистая линия — порог изменения фоновое качества для стимулов со спектром типа А. По оси абсцисс — частота, гц; по оси ординат — уровень звукового давления, дб от $2 \cdot 10^{-4}$ дин/см².

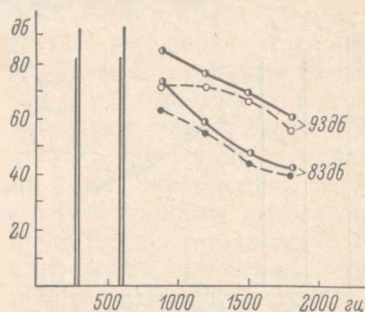


Рис. 5. Зависимость порога маскировки верхних гармоник (прерывистые линии) и фоновой границы (сплошные линии) от уровня интенсивности двух первых гармоник в спектре гласного (вертикальные линии).

Остальные обозначения те же, что на рис. 4.

гласном со спектром Б гармоника 900 гц лежит выше порога маскировки, но ниже порога изменения фоновое качества. Естественно было ожидать, что эта гармоника вызывает дополнительную маскировку более высоких составляющих.

Далее была сделана попытка создать такой сигнал из шести гармоник, чтобы гармоники 300 и 600 гц имели уровень интенсивности, равный 83 дб от $2 \cdot 10^{-4}$ дин/см², гармоника 900 гц была бы чуть ниже порога изменения фоновое качества, а гармоники 1200, 1500 и 1800 гц имели бы максимальные, но лежащие ниже порога маскировки значения. Процедура состояла в том, что комплекс, первоначально составленный из гармоник 300, 600 и 900 гц, постепенно наращивался прибавлением каждой следующей по номеру гармоники. Для каждой из этих гармоник сначала определялся порог ее маскировки комплексом, затем она сама включалась в комплекс и определялась маскировка

для следующей гармоник. Уровни составляющих в спектре полученного таким образом сигнала оказались несколько большими, чем в спектре *Б*. Это позволяет считать, что в отклике слуховой системы на «одноформантный» гласный со спектром *Б* также нет максимумов в области частот выше 900 гц.

Данные рис. 3 позволяют считать, что синтетические гласные со спектром *А* вполне могут быть аппроксимированы комплексами, состоящими всего из 3 гармоник с частотами F_0 , F_1 , F_2 . Все дальнейшие эксперименты были проведены с такими комплексами. Кроме того, совпадение порога маскировки с порогом изменения

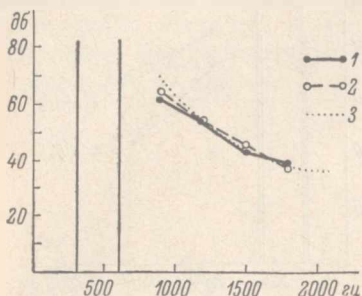


Рис. 6. Маскировка верхних гармоник 300 и 600 гц при длительности сигнала 200 (1) и 30 мсек. (2).

3 — порог изменения фонемного качества для стимулов со спектром типа *А*. Остальные обозначения те же, что на рис. 4.

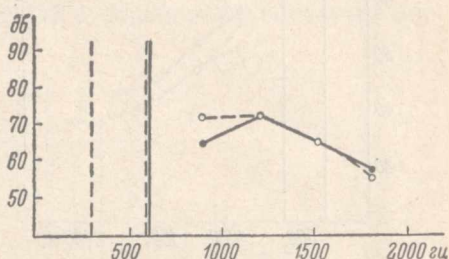


Рис. 7. Влияние гармоники 300 гц на порог маскировки верхних гармоник.

Сплошная линия — маскировка гармоникой 600 гц, прерывистая — маскировка гармониками 300 и 600 гц. Остальные обозначения те же, что на рис. 4.

фонемного качества позволило в дальнейшем отказаться от метода сравнения переменного сигнала с постоянным. Использование этого метода давало некоторое уменьшение разброса данных, однако это уменьшение оказалось существенным ($p < 0.001$) только у одного испытуемого.

Из рис. 3 следует, что человек способен обнаруживать дополнительные спектральные максимумы при очень малой их выраженности. Однако в наших опытах внимание испытуемых было сосредоточено на том, чтобы не пропустить момента перехода [о]-, [u]-образного гласного в другую фонему. Это можно рассматривать как придание другим фонемам ([i], [e], [ü], [ö]) большего веса, чем фонемам [о] и [u]. Для того чтобы выяснить, насколько это может быть существенным для результатов эксперимента, перед испытуемыми была поставлена следующая, в известном смысле обратная задача. Требовалось найти максимальное значение уровня гармоники, при котором еще можно принять гласный за [о] или [u].

Полученные данные приведены на рис. 4. Кривая зависимости максимального уровня гармоники от ее частоты близка по своей

форме к кривой маскировки, она только сдвинута вверх примерно на 7 дб. Рис. 5, где показан порог маскировки и положение фонемной границы, демонстрирует значение уровня интенсивности первой форманты. Фонемная граница определялась испытуемыми как такое значение стимула, при котором его с равным основанием можно интерпретировать как [o], [u] или как другую фонему ([a], [e], [ö], [i] или [ü]). Наблюдается хорошо известный в психоакустике факт изменения формы кривой маскировки с изменением уровня интенсивности маскира. Как показывает рис. 6, величина маскировки практически не меняется с укорочением стимула до 30 мсек.

Данные рис. 7 позволяют заключить, что первая гармоника (с частотой 300 гц) в нашем случае не оказывает влияния на величину маскировки на частотах 1200—1800 гц. Влияние этой гармоники на величину маскировки, составляющей 900 гц, легко объяснимо. Когда гармоника 300 гц отсутствует, добавление тона 900 гц к тону 600 гц воспринимается как понижение высоты звука, очевидно, связанное с появлением субъективного разностного тона с частотой 300 гц. Естественно, что, когда гармоника 300 гц сама присутствует в спектре стимула, этим критерием уже нельзя пользоваться: испытуемый должен ориентироваться на изменения в тембре гласного.

ЗАКЛЮЧЕНИЕ

Полученные данные подтверждают предположение, что для различения группы так называемых существенно одноформантных гласных ([u], [o]) от существенно двухформантных ([i], [e], [ö], [ü]) человек использует весьма простой признак — наличие или отсутствие максимума на слуховом эквиваленте спектра стимула в средней области частот. Об этом говорит тот факт, что пороговые по A_2 гласные действительно характеризуются присутствием дополнительного максимума на огибающей их спектра, а «одноформантный» синтетический гласный не отличается на слух от сигнала, не содержащего составляющих, а следовательно, и не имеющих его максимумов в данной частотной области.

Другое подтверждение того, что гласный описывается при восприятии указанием частот нескольких максимумов на слуховом эквиваленте спектра, было получено Хирато и др. (Hirato et al., 1968). Было показано, что для полной разборчивости гласных достаточно, чтобы гласный содержал всего 2—3 гармоники, расположенные в области частот его формант. Большая часть составляющих вообще может быть исключена из спектра гласного без того, чтобы человек мог это обнаружить.

Один из очень существенных вопросов, возникающих при разработке модели слухового обнаружения максимумов, касается выбора характеристик предварительных линейных фильтров, осуществляющих частотное разложение сигнала. Если исходить

из большинства психоакустических данных (см. обзор: Zwicker, Scharf, 1965; Чистович, 1966), добротность этих фильтров должна быть на порядок выше той, которая характеризует основную мембрану. Если бы добротность была мала, то максимальная составляющая гласного с частотой F_1 должна была бы сильно подавлять составляющую с частотой F_2 , этот эффект должен был бы медленно убывать с увеличением F_2/F_1 . Полученные в настоящей работе данные показывают, что эффект подавления сравнительно мал (составляющая с $F_2=2100$ гц обнаруживается, если она на 46 дБ ниже по уровню максимальной составляющей); эффект быстро убывает с увеличением расстояния по частоте. Нам не удалось найти в литературе данных для прямого сравнения с полученной нами зависимостью величины маскировки от расстояния по частоте между сигналом и маскером, однако похоже, что она принципиально не отличается от зависимостей, наблюдавшихся другими авторами при близких значениях частоты и интенсивности маскера (Чистович, 1957; Small, 1959; Green, 1965; Шупляков, 1967).

К настоящему времени накоплено уже довольно много данных о том, что слуховой частотный анализ речевых и речеподобных стимулов должен моделироваться набором фильтров с достаточно высокой добротностью (Plomp, 1964; Fujimura, 1967; Шупляков, 1967; Plomp, Mimpen, 1968).

Возникающее при этом затруднение состоит в том, что число обнаруживаемых слуховой системой максимумов на спектре гласного должно превышать число формант. Это в основном касается низкочастотной области спектра. Именно эта область кажется наиболее интересной для дальнейшего исследования восприятия амплитудных отношений между спектральными составляющими сигнала.

Л и т е р а т у р а

- В а р ш а в с к и й Л. А., И. М. Л и т в а к. 1955. Исследование формантного состава и некоторых других физических характеристик звуков русской речи. Пробл. физиол. акустики, III : 5—17.
- Ч и с т о в и ч Л. А. 1957. Частотные характеристики эффекта маскировки. Биофизика, 2 : 749—755.
- Ч и с т о в и ч Л. А. 1966. Психоакустика и вопросы теории восприятия речи. В сб.: Автоматическое распознавание слуховых образов, Новосибирск : 68—169.
- Ш у п л я к о в В. С. 1967. Слуховой анализ стационарных шумных согласных. Дисс. Л.
- F u j i m u r a O. 1967. On the second spectral peak of the front vowels: a perceptual study of the role of the second and third formants. Language a. Speech, 10, pt. 3 : 181—193.
- H i r a t o N., O. K a k u s h o, K. K a t o, T. K o b a y a s h i. 1968. Harmonic components and vowel perception. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 79—82.
- G r e e n D. 1965. Masking with two tones. J. Acoust. Soc. Amer., 37 : 802—813.

- Plomp R. 1964. The ear as a frequency analyzer. *J. Acoust. Soc. Amer.*, 36: 1628—1636.
- Plomp R., A. Mimpfen. 1968. The ear as a frequency analyzer, II. *J. Acoust. Soc. Amer.*, 43: 764—767.
- Small A. 1959. Pure-tone masking. *J. Acoust. Soc. Amer.*, 31: 1619—1625.
- Zwicker E., B. Scharf. 1965. A model of Loudness summation. *Psychol. Rev.*, 72: 3—26.
-

СЛУХОВОЕ ИЗМЕРЕНИЕ ПЕРВОЙ ФОРМАНТЫ

Л. А. Чистович, В. С. Шупляков

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

Данные по восприятию синтетических гласных (Hirato et al., 1968; Lindqvist, Pauli, 1968; Мушников, Чистович, 1971a; Слепокурова, 1971) показывают, что связь между акустическими характеристиками гласного и его фонемным обозначением довольно просто описывается, если в качестве параметров сигнала использовать частоты формант. Это позволяет думать, что выделяемые слухом признаки гласного или совпадают с формантами, или являются достаточно близкими к ним.

В технике известен ряд методов приближенного измерения частот формант (Сапожков, 1963; Flanagan, 1965). Один из них, основанный на выделении максимумов на кривой отклика гребенки предварительных частотных фильтров, иногда рассматривается как аналог слухового определения формант (Campanella, Phylfe, 1967).

Так как факт предварительной фильтрации сигнала в механических системах улитки не вызывает сомнений, а наличие механизмов «латерального торможения», способных обеспечить выделение максимумов, достаточно вероятно, эта гипотеза представляется правдоподобной и заслуживающей проверки.

Один из наиболее сложных вопросов, возникающих при попытках представить данную гипотезу в форме модели, касается выбора характеристик предварительных фильтров. Если принять характеристики, определенные Бекеша для базилярной мембраны (Békésy, 1960), взаимная маскировка формант оказывается в модели значительно большей, чем у человека (Мушников, Чистович, 1971b; Лабутин, Павловский, 1971).

Если же принять характеристики фильтров, согласующиеся с психоакустическими данными по маскировке (Темов, 1971), появляются трудности в определении первой форманты. В области частот ниже 500—600 гц ширина критической полосы меньше 100 гц (Zwicker et al., 1957; Plomp, 1964; Plomp, Mimpfen, 1968); следовательно, даже при низких частотах основного тона макси-

мумы на отклике модели должны соответствовать не форманте (резонансной частоте фильтрующей цепи речевого тракта или синтезатора), а отдельным гармоническим составляющим гласного. Изменение F_1 гласного будет отражаться в модели в виде изменения относительной выраженности максимумов, соответствующих гармоникам. Если в качестве представляющего F_1 принимается наиболее выраженный максимум, то результатом плавного изменения стимула будут скачкообразные перемещения «воспринимаемой» частоты F_1 с одной гармоники на другую.

Описываемые ниже эксперименты были предприняты для выяснения того, не наблюдается ли подобный эффект при слуховом восприятии первой форманты.

Основой для выбора методики послужило следующее наблюдение: было замечено, что ряд [i]—[e]—[æ] гласных может быть синтезирован путем суммирования выходного сигнала фильтра F_2 параллельного формантного синтезатора с чистым тоном низкой частоты. Изменение частоты тона вызывает два эффекта: изменение высоты гласного (это явление подробно не исследовалось) и изменение фонетического качества. Фонетическое качество сигналов настолько отчетливо, что испытуемые без труда могут определять фонемные границы ([i]—[e] и [e]—[æ]) и сравнивать синтезированные таким способом гласные с естественными и синтетическими двухформантными гласными. Никаких неприятных эффектов типа хриплости или биений (их можно было ожидать из-за некогерентности тона и основной частоты возбуждения второй форманты) при этом не наблюдается, если частота второй форманты существенно выше частоты тона.

Это наблюдение дало возможность провести эксперименты по сравнению двухформантных синтетических гласных с гласными, где первая форманта была заменена чистым тоном. Испытуемый, управляя частотой тона, должен был установить такое его значение, при котором оба гласных оказываются наиболее близкими (совпадающими) по фонетическому качеству (на высоту звука испытуемый не должен был обращать внимание). Частота тона в таком совпадающем по фонетическому качеству гласном принималась за воспринимаемую частоту F_1 двухформантного гласного.

МЕТОДИКА

Для экспериментов была использована установка, блок-схема которой приведена на рис. 1. С генератора основного тона сигнал подавался на фильтрующие цепи, резонансные частоты которых находились в области первой F_1 и второй F_2 формант гласных. С выхода сумматора I подавался сигнал на телефон испытуемого для прослушивания. Сигнал второй форманты подавался, кроме того, на сумматор II, где он смешивался с тоном от звукового генератора и также подавался на телефон. Таким образом, испытуемый имел возможность поочередно прослушивать сигналы, отличающиеся только видом первой форманты: в одном случае это был набор гармонических составляющих, с огибающей, имеющей форму резонансной кривой избирательного эле-

мента, а в другом — одну-единственную составляющую. Атенюаторы использовались для предварительного выравнивания получаемых звуков по громкости. Задача испытуемого заключалась в том, чтобы при различных значениях F_1 подобрать такие значения частоты f звукового генератора, при которых получаемые сигналы были наиболее похожи по фонетическому ка-

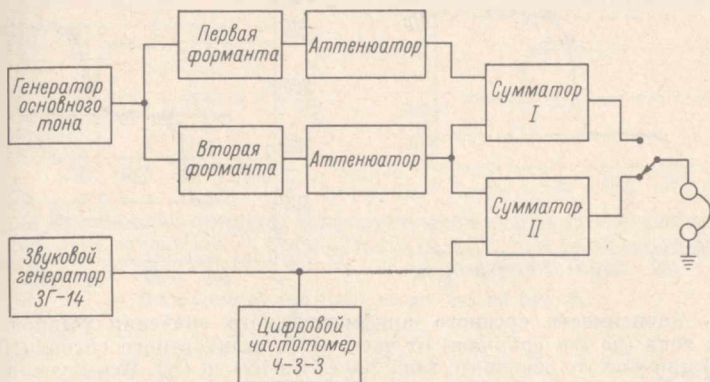


Рис. 1. Блок-схема установки для проведения экспериментов.

Описание в тексте.

честву. Частота звукового генератора затем изменялась цифровым частотомером. Значение частоты второй форманты сохранялось одинаковым во всех экспериментах и было равно 2300 гц. Частота первой форманты F_1 изменялась от 300 до 600 гц скачками через 20 гц. Ширина первой форманты, измеренная на уровне 3 дб, была равна 70 гц. На каждой частоте испытуемый производил по 5 или 10 измерений. Эксперименты проводились при различных частотах основного тона, принимающих значения от 80 до 200 гц.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В предварительных измерениях, сделанных одним испытуемым, применялись гласные с частотами основного тона 130, 150 и 200 гц. Каждое измерение состояло из 5 повторных установок частоты тона.

Полученные зависимости среднего арифметического значения установленной частоты тона в переменном сигнале от частоты F_1 стандартного сигнала приведены на рис. 2, который показывает, что испытуемый устанавливает частоту тона на одну из гармоник в стандартном сигнале, расположенных в области его первой форманты, а не на резонансную частоту (форманту) избирательного элемента. Об этом же свидетельствует гистограмма значений отдельных установок частоты (рис. 3), обнаруживающая отчетливые пики на частотах гармоник.

Основные эксперименты, в которых участвовало трое испытуемых, были проведены на гласных с низкими частотами основного тона: 80 и 100 гц — испытуемые В. Ш. и Е. Т.; 100 и 120 гц — испытуемая Л. Ч. Каждое измерение состояло из 10 повторных установок частоты тона.

Полученные результаты приведены на рис. 4—6. Можно видеть, что у В. Ш. и Е. Т. при $F_0=100$ гц зависимость имеет еще

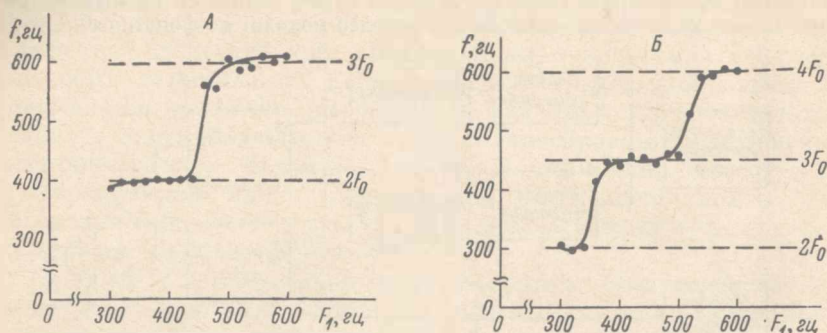


Рис. 2. Зависимость среднего арифметического значения установленной частоты тона (по оси ординат) от частоты F_1 стандартного сигнала (по оси абсцисс) при частоте основного тона 200 гц (А) и 150 гц (Б). Испытуемая Л. Ч.

Прерывистые линии — значения частоты тона, соответствующие гармоникам F_0 в спектре стандартного сигнала.

отчетливо ступенчатый характер, совпадение установленной частоты тона с гармониками в стандартном сигнале достаточно

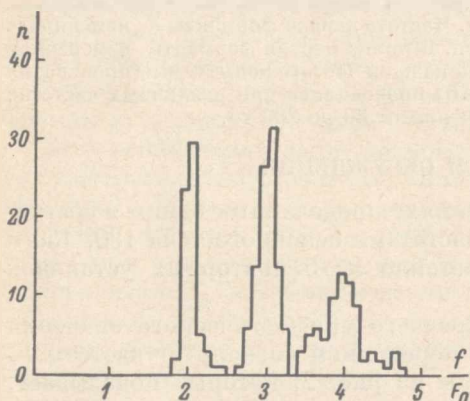


Рис. 3. Гистограмма значений отдельных установок частоты. Испытуемая Л. Ч.

Частоты основного тона 130, 150 и 200 гц. По оси абсцисс — отношение частоты тона к основной частоте стандартного сигнала; по оси ординат — число случаев. Общее число установок частоты $N=240$.

очевидно. Только при $F_0=80$ гц (и $F_0=100$ гц у Л. Ч.) кривая зависимости приобретает плавный характер, позволяющий предполагать, что в данном случае измеряются уже не гармоники, а некоторый суммарный спектральный максимум, соответствующий форманте.

Однако плавный характер кривой может быть и результатом усреднения данных при обработке. Если при данном значении F_1 стандартного сигнала испытуемый в части случаев устанавливает частоту тона на ближайшую снизу, а в части случаев — на

ближайшую сверху гармонику, естественно, что среднее арифметическое из установленных значений частоты тона будет находиться где-то в промежутке между частотами гармоник. В случае справедливости этого последнего предположения должны, од-

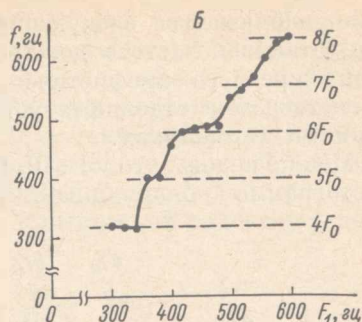
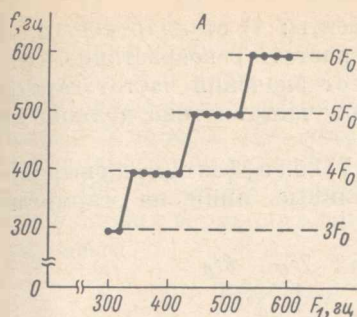


Рис. 4. Зависимость среднего арифметического значения установленной частоты тона от частоты F_1 стандартного сигнала при частоте основного тона 100 (А) и 80 гц (Б). Испытуемый В. Ш.

Остальные обозначения те же, что на рис. 2.

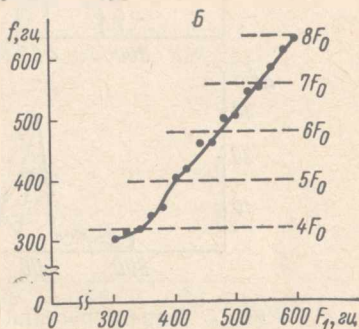
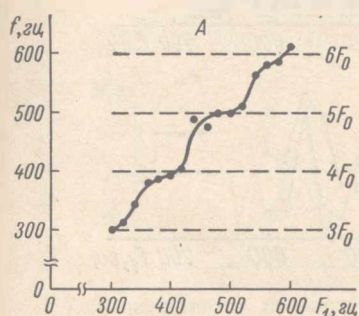


Рис. 5. Зависимость среднего арифметического значения установленной частоты тона от частоты F_1 стандартного сигнала при частоте основного тона 100 (А) и 80 гц (Б). Испытуемая Е. Т.

Остальные обозначения те же, что на рис. 2.

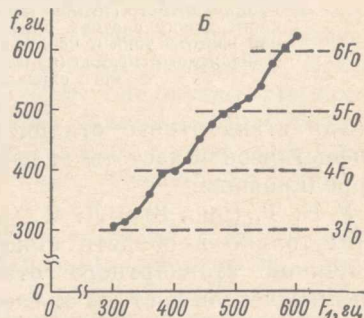
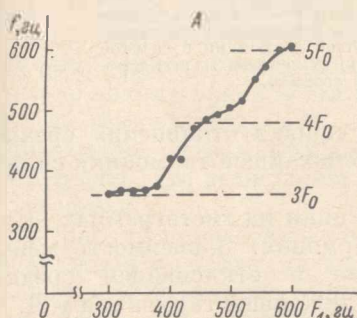


Рис. 6. Зависимость среднего арифметического значения установленной частоты тона от частоты F_1 стандартного сигнала при частоте основного тона 120 (А) и 100 гц (Б). Испытуемая Л. Ч.

Остальные обозначения те же, что на рис. 2.

нако, наблюдаться следующие эффекты: 1) стандартное отклонение установок частоты должно возрастать с возрастанием отклонения среднего арифметического от значений частот гармоник; 2) гистограмма отдельных установок частоты тона должна иметь пики на гармониках.

Можно видеть, что у В. Ш. (рис. 7) оба эффекта ярко выражены. Гистограмма обнаруживает отчетливые пики на гармониках,

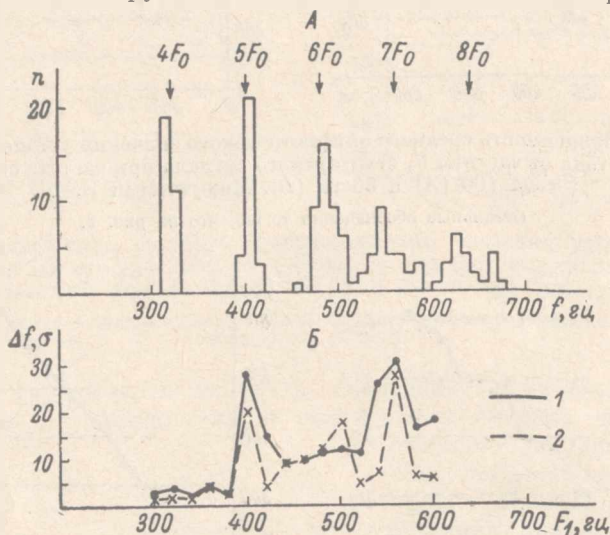


Рис. 7. Характеристики совокупности установленных значений частоты тона. Испытуемый В. Ш.

Частота основного тона 80 гц. А — гистограмма значений отдельных установок частоты тона (по оси абсцисс — частота тона, по оси ординат — число случаев; стрелками указаны значения частоты тона, соответствующие гармоникам F_n в спектре стандартного сигнала); на Б: 1 — зависимость стандартного отклонения установок частоты тона (по оси ординат) от частоты F_1 стандартного сигнала (по оси абсцисс), 2 — зависимость разности между средним арифметическим значением установленной частоты тона и частотой ближайшей гармоники в спектре стандартного сигнала (по оси ординат) от частоты F_1 стандартного сигнала (по оси абсцисс).

кривая стандартного отклонения и кривая отклонения средней арифметической частоты от частоты ближайшей гармоники сильно коррелированы.

У Е. Т. (рис. 8) и Л. Ч. (рис. 9) пики на гистограммах выражены только в области нижних гармоник. Зависимость между величиной стандартного отклонения и отклонением средней арифметической частоты от частоты ближайшей гармоники у Л. Ч. прослеживается только в низкочастотной области, у Е. Т. ее вообще не видно.

Таким образом, можно сделать вывод, что у В. Ш. и при $F_0 = 80$ гц не происходит объединения гармоник в общий спектраль-

ный максимум. У Е. Т. и Л. Ч. объединения нет, если форманта расположена в нижней части исследованного частотного диапазона, и оно, возможно, есть, когда частота форманты выше 500 гц.

Полученные результаты не противоречат данным по измерению критической полосы. На частоте 500 гц ширина критической полосы составляет примерно 80—120 гц (Zwicker et al., 1957; Plomp, Mimpen, 1968). Данные Пломпа и Мимпена свидетельствуют о довольно больших вариациях ширины критической полосы у разных испытуемых.

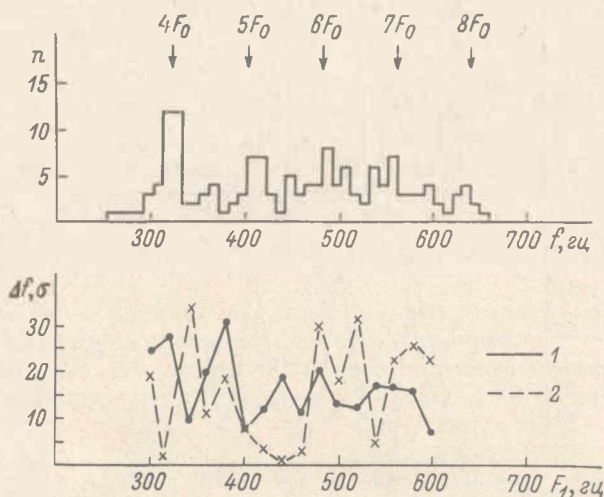


Рис. 8. Характеристики совокупности установленных значений частоты тона. Испытуемая Е. Т.

Частота основного тона 80 гц. Остальные обозначения те же, что на рис. 7.

До сих пор полученные факты рассматривались только с той точки зрения, измеряет ли человек частоту форманты или частоты гармоник, расположенных вблизи частоты форманты. Вопрос о том, по какому правилу выбирается одна, а не другая гармоника, пока не обсуждался. Хотя полученные данные не содержат исчерпывающей информации для ответа на этот вопрос, они позволяют думать, что выбирается просто более интенсивная (громкая) из гармоник.

На рис. 10 приведены данные для гласных с $F_0=100$ гц. Горизонтальными прямыми показаны диапазоны значений F_1 , внутри которых испытуемые не менее чем в 9 случаях из 10 устанавливали частоту тона, более близкую к гармонике указанного номера, чем к любой другой. Стрелками показаны значения F_1 , соответствующие равной вероятности выбора каждой из двух соседних гармоник.

Можно видеть, что равная вероятность выбора гармоник 500 и 600 гц приходится на $F_1 \approx 550$ гц, т. е. на такую частоту форманты, при которой обе эти гармоники должны быть примерно одинаковыми по амплитуде. Точки равной вероятности выбора гармоник 400 и 500 гц и гармоник 300 и 400 гц несколько смещены вниз от F_1 , равной 450 гц и 350 гц соответственно. Это может быть связано с влиянием частотной характеристики слуха (при равной

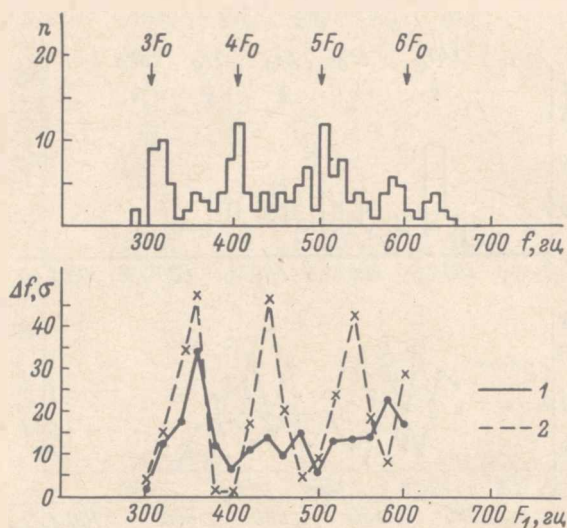


Рис. 9. Характеристики совокупности установленных значений частоты тона. Испытуемая Л. Ч.

Частота основного тона 100 гц. Остальные обозначения те же, что на рис. 7.

интенсивности гармоник 300 и 400 гц вторая из них является более громкой).

Как уже говорилось, обработка данных включала вычисление стандартных отклонений установки частоты (для случаев, когда измерение состояло из 10 повторных установок). На рис. 11 приведена гистограмма значений стандартных отклонений, выраженных в % от среднего арифметического значения. Медиана распределения соответствует 2.5%. Это довольно близко к величинам дифференциального порога для второй форманты (Flanagan, 1955; Kakusho et al., 1968). Заманчиво предположить, что эта величина отражает точность пространственного механизма измерения частоты, т. е. точность локализации максимума на отклике гребенки фильтров.

ЗАКЛЮЧЕНИЕ

Полученные данные показали, что человек не производит измерения частоты первой форманты как таковой (т. е. как резонансной частоты фильтрующей цепи речевого тракта), а поль-

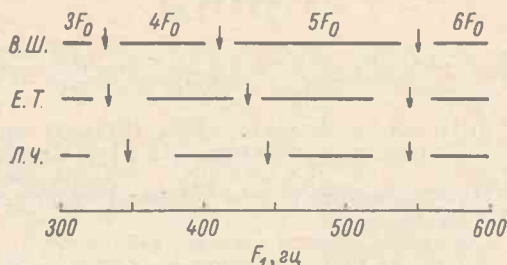


Рис. 10. Связь между номером выбранной гармоники в спектре стандартного сигнала и частотой F_1 стандартного сигнала.

Объяснения в тексте.

зуется ее приближенной оценкой, ориентируясь на частоту наиболее выраженной спектральной составляющей.

Это значит, что отпадает гипотеза слухового анализа через синтез (Stevens, 1960; Nakatsui, Suzuki, 1968), которая, как известно, была выдвинута, для того чтобы объяснить возможность точного вычисления значения частоты форманты по спектральной картине сигнала.

Результаты экспериментов согласуются с психоакустическими данными о разрешающей способности слуховой системы как анализатора частот.

Кажется целесообразным продолжать работу по созданию модели выделения максимумов, приняв в качестве характеристик филь-

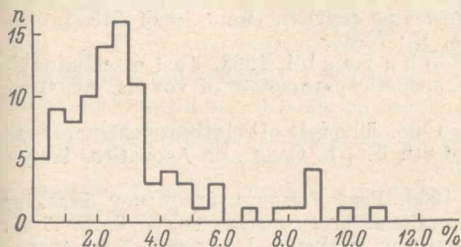


Рис. 11. Гистограмма значений стандартных отклонений установок частоты тона.

По оси абсцисс — значение стандартного отклонения, в % от среднего арифметического значения установленной частоты тона; по оси ординат — число случаев.

ров кривые маскировки, определенные в психоакустическом эксперименте. Так как наиболее серьезной проверкой модели будет ее поведение при действии нестационарных сигналов, необходимо найти психоакустические методы исследования восприятия формант в изменяющихся во времени гласных. Первым шагом в этом направлении может быть проведение экспериментов,

подобных описанным выше, с использованием в качестве стандартного сигнала синтетического гласного с изменяющейся во времени основной частотой.

Л и т е р а т у р а

- Л а б у т и н В. К., В. В. П а в л о в с к и й. 1971. Изучение реакции основной мембраны улитки на сложные акустические стимулы. Наст. сб. : 29—36.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971а. О слуховом описании гласного. I. Признаки, различающие [i] и [e]. Наст. сб. : 5—10.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971б. О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном. Наст. сб. : 11—19.
- С а п о ж к о в М. А. 1963. Речевой сигнал в кибернетике и связи. М.
- С л е п о к у р о в а Н. А. 1971. О положении формантной границы между синтетическими гласными [i]—[e], [ö]—[ü] и [o]—[u]. Наст. сб. : 136—142.
- Т е м о в В. Л. 1971. Модель для описания результатов психоакустических экспериментов со стационарными сигналами. Наст. сб. : 36—49.
- B é k é s y G. 1960. Experiments in hearing. N. Y.—Toronto—London.
- C a m p a n e l l a J., D. P h y f e. 1967. Application of a model of the analog ear to speech signal analysis. Confer. on Speech Communication and Processing, Boston : 198—206.
- F l a n a g a n J. L. 1955. A difference Limen for Vowel Formant Frequency. J. Acoust. Soc. Amer., 27 : 613—617.
- F l a n a g a n J. L. 1965. Speech analysis, synthesis and perception. Berlin—Heidelberg—N. Y.
- H i r a t o N., O. K a k u s h o, K. K a t o, T. K o b a y a s h i. 1968. Harmonic components and vowel perception. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 79-82.
- K a k u s h o O., K. K a t o, T. K o b a y a s h i. 1968. Just discriminable change and matching range of acoustic parameters of vowels. Acustica, 20, 1 : 46—54.
- L i n d q v i s t J. Ch., S. P a u l i. 1968. The role of relative spectrum levels in vowel perception. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 91—94.
- N a k a t s u i M., J. S u z u k i. 1968. Fast Formant Frequency Tracking Technique. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, B : 47-50.
- P l o m p R. 1964. The ear as a frequency analyzer. J. Acoust. Soc. Amer., 36 : 1628—1636.
- P l o m p R., A. M. M i m p e n. 1968. The ear as a frequency analyzer II. J. Acoust. Soc. Amer., 43 : 764—767.
- S t e v e n s K. N. 1960. Toward a model for speech recognition. J. Acoust. Soc. Amer., 32, 1 : 47—55.
- Z w i c k e r E., E. F l o t t o r p, S. S. S t e v e n s. 1957. Critical bandwidth in loudness summation. J. Acoust. Soc. Amer., 29, 5 : 548—557.

ИЗУЧЕНИЕ РЕАКЦИИ ОСНОВНОЙ МЕМБРАНЫ УЛИТКИ НА СЛОЖНЫЕ АКУСТИЧЕСКИЕ СТИМУЛЫ

В. К. Лабутин, В. В. Павловский

Ленинград

Результаты известных исследований Бекеши (Békésy, 1960) и их теоретическая аппроксимация (Zwislocki, 1950; Flanagan, 1960) составляют фундамент современных представлений о частотно-селективных свойствах периферического отдела слухового анализатора. Они дают описание частотных и координатных характеристик основной мембраны (стационарный отклик на чистый тон), а также используются для определения ее переходных характеристик (Bauch, 1956). При сопоставлении этих данных с результатами психоакустических исследований дифференциальных порогов по частоте и данными по маскировке чистых тонов (Чистович, 1957; Chocholle, 1955), а также с данными электрофизиологических исследований частотно-пороговых кривых нейронов (Гершуни, 1967; Katsuki et al., 1958) возникают серьезные трудности, для преодоления которых выдвигаются многочисленные гипотезы о механизмах обострения частотно-избирательных характеристик основной мембраны (Huggins, Licklider, 1951; Любинский, Позин, 1967).

Из теории колебаний и экспериментов, проводимых на электронных моделях различных механизмов обострения (Лабутин, Молчанов, 1967), вытекает, что при одновременном воздействии двух или большего числа колебаний на нелинейную систему неизбежны специфические эффекты типа маскировки (подавление слабого сигнала сильным), разностного тона (появление колебаний разностной частоты) и другие, ухудшающие различимость сигналов в пределах полосы частот, пропускаемой предшествующим нелинейной цепи линейным частотно-избирательным фильтром. Это обстоятельство существенно ограничивает эффективность нелинейных (реализуемых в моделях нейронных цепей) механизмов обострения в приложении к задачам разделения монохроматических стимулов, каковыми по существу являются все естественные акустические сигналы, воспринимаемые органом слуха.

Таким образом, для дальнейшего изучения механизмов частотного анализа сигналов органом слуха особый интерес представляет исследование реакции слухового анализатора на сложные сигналы. В плане таких исследований было предпринято совместное психоакустическое и модельное изучение реакций, вызываемых широкополосными сигналами типа синтетических гласных.

В результате психоакустического исследования (Мушников, Чистович, 1971) был найден набор синтетических двухформантных гласных, соответствующий порогу обнаружения человеком наличия в этих гласных второй форманты. Спектры этих сигналов,

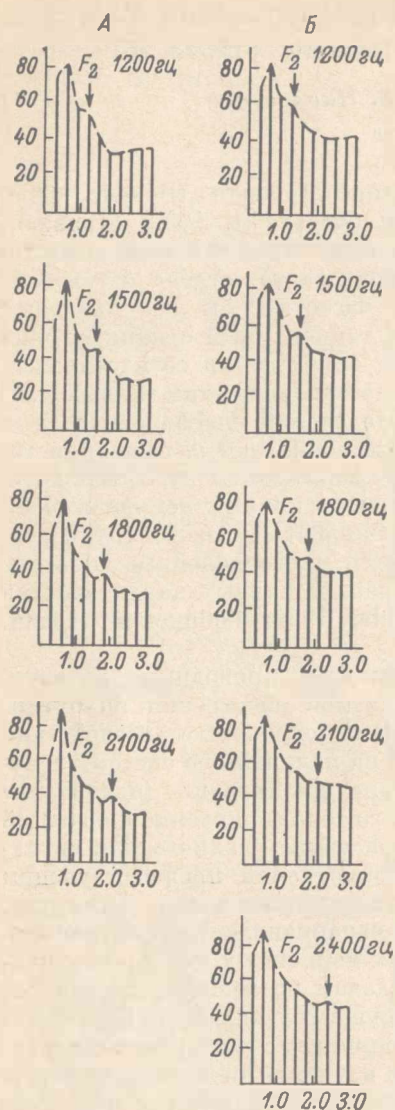


Рис. 1. Спектры синтетических сигналов на уровне звукового давления.

По оси абсцисс — частота, кгц; по оси ординат — уровень звукового давления, дб от $2 \cdot 10^{-4}$ дин/см². Для сигналов группы А наклон огибающей спектра основного тона — 12 дб/окт.; для сигналов группы Б — 6 дб/окт.

с учетом частотной характеристики телефона, приведены на рис. 1. Можно видеть, что эти спектры характеризуются наличием на огибающей двух максимумов: сильно выраженного максимума, соответствующего первой форманте (600 гц), и слабого дополнительного максимума, соответствующего второй форманте.

Целью модельного исследования было выяснение того, выделяется ли этот слабый максимум на отклике модели улитки, или же он остается замаскированным. В качестве характеристики сигналов, использованных при описываемых ниже расчетах, взяты значения их спектров на уровне электрического напряжения, подводимого к телефону. Они приведены в табл. 1. В табл. 2 даны значения спектров на уровне звукового давления.

Отклик k -й точки основной мембраны определялся «энергетическим» суммированием откликов этой точки на все частотные составляющие сигнала:

$$d_k = \sqrt{\sum_i (A_i G_i y_{ik})^2}, \quad (1)$$

где A_i — напряжение i -й составляющей, подведенное к телефону; G_i — коэффициент передачи телефона и структур, предшествующих мембране овального окошка; y_{ik} — ордината амплитудно-частотной характеристики k -й точки основной мембраны на частоте i -й составляющей.

Коэффициент G_i принят состоящим из двух частотно-зависимых сомножителей:

$$G_i = K_{\tau i} K_{\text{ср} i}, \quad (2)$$

где $K_{\tau i}$ — отдача телефона на частоте i -й составляющей; $K_{\text{ср} i}$ —

отношение амплитуды колебаний мембраны овального окошка к давлению на барабанной перепонке. Коэффициент передачи давления в наружном ухе принят частотно независимым. Значения $K_{ср}$, взяты из экспериментальных данных Бекеши (рис. 2, сплошная линия), экстраполированных до частоты 3 кгц (штриховая линия).

Расчет по формуле (1) произведен для 60 точек основной мембраны, охватывающих диапазон координатных частот от 150 до 3000 гц с одинаковым относительным интервалом частот:

$$\frac{f_{k+1}}{f_k} = \sqrt[60]{\frac{3000}{150}} = 1.05208. \quad (3)$$

В качестве амплитудно-частотных характеристик точек основной мембраны y_k использована предложенная Фланаганом (Flanagan, 1960) аналитическая аппроксимация данных Бекеши:

$$F_1(j\xi) = C_1 \beta^r \frac{\frac{\epsilon}{\beta} + j\xi}{\frac{\gamma}{\beta} + j\xi} \times e^{-j\xi\beta T} \times \frac{1}{\left[\left(1 + \frac{1}{k^2} - \xi^2 \right) + j \frac{2}{k} \xi \right]^2}, \quad (4)$$

где $\epsilon/\beta=0.1$; $\gamma/\beta=1$; $k=2$; β — координатная частота данной точки основной мембраны; ξ — относительная частота возбуждения (f/β); r и T — константы.

Модуль коэффициента передачи (4) равен

$$|F_1(\xi)| = \sqrt{\frac{\xi^2 + 0.01}{\xi^2 + 1}} \frac{c_1 \beta^r}{\xi^4 - 1.5\xi^2 + 1.562} \quad (5)$$

и, переходя к нашим обозначениям, получим:

$$y_{ik} = \sqrt{\frac{(f_i/f_k)^2 + 0.01}{(f_i/f_k)^2 + 1}} \frac{1.4942\varphi(k)}{(f_i/f_k)^n - 1.5(f_i/f_k)^2 + 1.562}, \quad (6)$$

где $\varphi(k)$ отражает ход максимумов амплитудно-частотных характеристик различных точек основной мембраны, а $n=4$.

По Фланагану (Flanagan, 1960),

$$\varphi(k) = a(2\pi f_k)^{0.83} \quad (7)$$

и удовлетворительно аппроксимирует экспериментальные данные Бекеши только до частоты 1 кгц (рис. 3). Поэтому в наших расче-

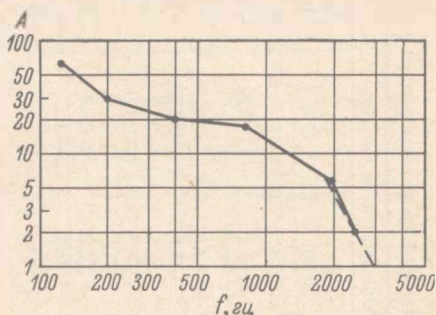


Рис. 2. Зависимость амплитуды колебаний мембраны овального окошка (в ангстремах) от частоты при постоянной амплитуде давления на барабанной перепонке (1 дин/см²) (по Békésy, 1960).

Т а б л и ц а 1

Спектры синтетических сигналов на уровне электрического напряжения (в дБ относительно уровня второй гармоники)

Частота гармоники, Гц	Сигнал и частота второй форманты F_2 , Гц								
	A1 1200	A2 1500	A3 1800	A4 2100	B1 1200	B2 1500	B3 1800	B4 2100	B5 2400
300	-23.5	-24.5	-24.8	-24.5	-13.1	-13.6	-13.4	-13.4	-13.6
600	0	0	0	0	0	0	0	0	0
900	-25.4	-25.7	-25.7	-25.4	-15.4	-14.4	-15.0	-14.4	-14.2
1200	-24.2	-33.6	-33.0	-33.0	-16.9	-22.2	-22.0	-21.8	-21.5
1500	-36.4	-31.4	-40.0	-38.2	-24.1	-22.7	-26.4	-26.0	-25.7
1800	-42.4	-40.0	-36.8	-42.6	-27.9	-29.0	-25.2	-30.2	-29.7
2100	-41.6	-45.1	-44.0	-38.2	-30.5	-31.6	-31.2	-27.4	-32.7
2400	-41.2	-46.3	-45.6	-46.3	-33.2	-34.1	-33.8	-33.2	-31.2
2700	-43.2	-52.3	-51.7	-51.6	-35.3	-36.4	-36.0	-33.7	-35.5
3000	-44.2	-52.3	-52.0	-51.6	-36.5	-37.8	-38.6	-38.2	-38.3

Т а б л и ц а 2

Спектры синтетических сигналов на уровне звукового давления (в дБ относительно уровня второй гармоники)

Частота гармоники, Гц	Сигнал и частота второй форманты F_2 , Гц									
	A1 1200	A2 1500	A3 1800	A4 2100	B1 1200	B2 1500	B3 1800	B4 2100	B5 2400	
300	-23.5	-24.5	-24.8	-24.5	-13.1	-13.6	-13.4	-13.4	-13.6	
600	0	0	0	0	0	0	0	0	0	
900	-27.4	-27.7	-27.7	-27.4	-17.4	-16.4	-17.0	-16.4	-16.2	
1200	-28.7	-38.1	-37.5	-37.5	-21.4	-26.7	-26.5	-26.3	-26.0	
1500	-42.8	-37.8	-46.4	-44.6	-30.5	-29.1	-32.8	-32.4	-32.1	
1800	-50.4	-47.0	-44.8	-50.6	-35.9	-37.0	-33.2	-38.2	-37.7	
2100	-50.0	-53.5	-52.4	-46.6	-38.9	-40.0	-39.6	-41.1	-39.4	
2400	-48.4	-54.5	-52.8	-53.5	-40.4	-41.3	-41.0	-40.4	-38.4	
2700	-47.2	-56.3	-55.7	-55.6	-39.3	-40.4	-40.0	-39.7	-39.5	
3000	-44.2	-54.3	-54.0	-53.6	-38.5	-39.8	-40.6	-40.2	-40.3	

тах применена смешанная аппроксимация значений $\varphi(k)$: до частоты 1 кГц — по формуле (7), а на частотах выше 1 кГц принято постоянное значение, лучше соответствующее данным Бекеша (рис. 4).

Ввиду появления целого ряда указаний на наличие механизмов, обеспечивающих весьма крутой спад частотных характеристик в сторону верхних частот (Чистович, 1957; Katsuki et al., 1958), расчеты выполнены также со значением показателя $n=16$, что соответствует правому склону амплитудно-частотной характеристики с асимптотой 96 дБ/окт.

Учитывая широкое использование биоанализаторами механизмов сравнения, помимо вычисления смещений точек основной

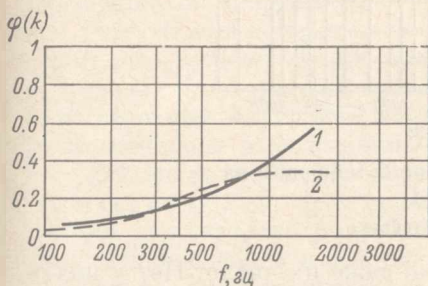


Рис. 3. Сопоставление частотных характеристик коэффициента $\varphi(k)$ по Фланегану (1) и по Бекешу (2).

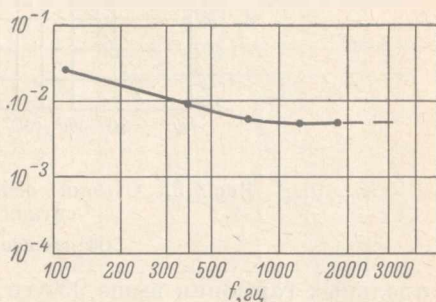


Рис. 4. Отношение максимального смещения основной мембраны к амплитуде колебаний стремечка в зависимости от частоты (по Békésy, 1960).

мембраны по формуле (1), мы определили первые разности смещений соседних точек основной мембраны:

$$\Delta d_k = d_{k+1} - d_k. \quad (8)$$

Вычисление этих разностей моделирует элементарную операцию широко обсуждаемых в литературе нелинейных механизмов обострения частотной селекции в нейронной части слуховой системы.

Все расчеты выполнены на ЦВМ. Их результаты показывают отсутствие заметных отличий как в огибающих смещений точек основной мембраны, так и в их первых разностях. В качестве примера на рис. 5 и 6 приведены результаты расчета откликов основной мембраны на сигналы с наибольшими различиями в спектрах (сигналы А1 и А4 в табл. 1). Сплошными линиями нанесены кривые, полученные при $n=4$ в формуле (6), штриховыми — при $n=16$. Вертикальными отрезками прямых на этих же графиках представлены спектры сигналов (напряжения, подведенного к телефону).

Как видно из рис. 5 и 6, в структуре кривых $d_k = (f_k)$ и $\Delta d_k = (f_k)$ отсутствуют признаки, отражающие не только величину

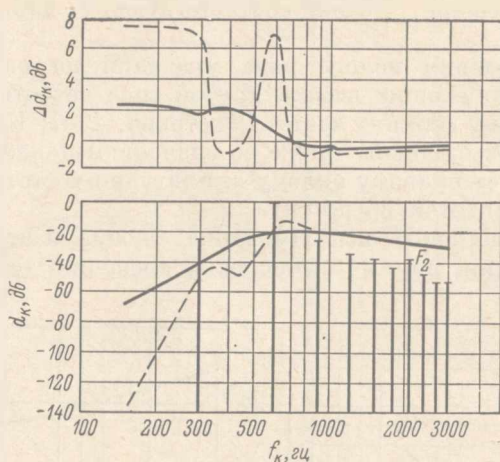


Рис. 5. Отклик основной мембраны на сигнал А1.

Объяснение в тексте.

отдельных гармоник выше 900 Гц, но даже их число. Нет у полученных кривых и каких-либо отличий, позволяющих судить

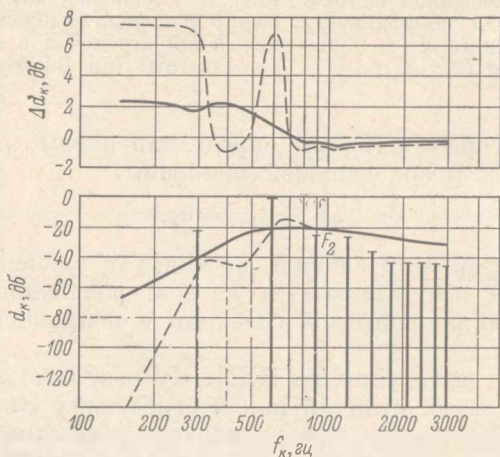


Рис. 6. Отклик основной мембраны на сигнал А4.

Объяснение в тексте.

о значении частоты второй форманты. Различия между кривыми откликов на разные сигналы не превышают 0.1—0.2 дБ по уровню и 0.1—0.2% по координатной частоте, когда $n=4$, т. е. эти раз-

личия ниже дифференциальных порогов слуха. При значении $n=16$ различия возрастают и достигают известных значений дифференциальных порогов слуха: порядка 0.5 дБ по уровню и 1% по частоте. Дополнительные расчеты высших порядков разностей при $n=4$ (рис. 7) не позволили выявить структуру спектра сигнала в области второй форманты и привели к быстрому росту шума, несмотря на высокую точность расчетов на ЦВМ (до шестого знака) и плавный вид аналитических аппроксимаций. Единственный излом (в аппроксимации функции $\varphi(\kappa)$ на частоте 1 кГц — рис. 4) привел к четко выраженным разрывам в ходе зависимостей высших разностей от координаты.

Полученные данные показывают, что существующих представлений о механизмах частотной селекции на уровне основной мембраны в сочетании со средствами нелинейной обработки снимаемой с нее информации недостаточно для объяснения наблюдаемой в психоакустических экспериментах разрешающей способности слуха. Представляется весьма вероятным, что предшествующие механизму спайковой активности рецепторных нейронов линейные частотные фильтры обладают

значительно более острыми амплитудно-частотными характеристиками, чем известные характеристики основной мембраны, и имеют полосы пропускания, приближающиеся к критическим. Для уточнения имеющихся представлений большую ценность представили бы дополнительные исследования колебательных характеристик улитки, в особенности ее базальной части, ответственной за восприятие частот выше 1 кГц, и механизмов возбуждения волосковых и биполярных клеток. Заслуживают внимания также гипотезы о линейных механизмах обострения частотной селекции основной мембраны и иных возможностях линейной селекции на периферии слухового анализатора. В этой связи уместно отметить появившуюся при подготовке настоящей статьи публикацию, подтверждающую существование линейных механизмов, которые обеспечивают более высокую частотную избирательность по сравнению с обнаруженной на уровне основной мембраны в опытах Бекеша (de Boer, Jongkees, 1968).

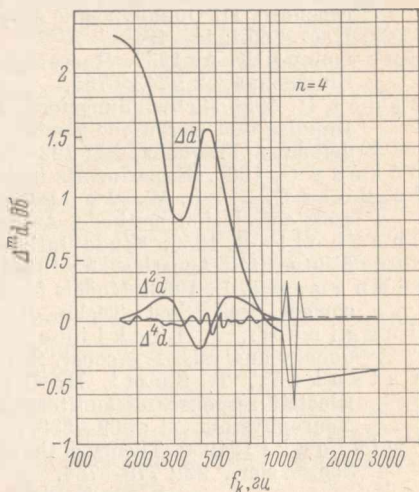


Рис. 7. Результаты расчета разностей высших порядков.

Объяснение в тексте.

Л и т е р а т у р а

- Г е р ш у н и Г. В. 1967. О механизмах слуха (в связи с исследованием временных и временно-частотных характеристик слуховой системы). В сб.: Механизмы слуха, М.—Л. : 3—32.
- Л а б у т и н В. К., А. П. М о л ч а н о в. 1967. Слух и анализ сигналов. М.—Л.
- Л ю б и с к и й И. А., Н. В. П о з и н. 1967. Анализ модели однородного нейронного слоя с латеральными связями. Автоматика и телемеханика, 10 : 168—181.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971. О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном. Наст. сб. : 11—19.
- Ч и с т о в и ч Л. А. 1957. Частотные характеристики эффекта маскировки. Биофизика, 2 : 743—755.
- B a u c h H. 1956. Schwindungsform der Basilmembran bei Erregung durch Impulse und Geräusche gemessen an einem elektrischen Modelle der Innenohres. Frequenz, 10 : 222—234.
- B é k é s y G. 1960. Experiments in hearing. N. Y.
- de B o e r E., L. B. W. J o n g k e e s. 1968. On cochlear sharpening and cross-correlation methods. Acta Oto-Laryngol., 65 : 97—104.
- C h o c h o l l e R. 1955. Etude statistique de la sensibilité auditive différentielle d'intensité. Acustica, 5 : 134—141.
- F l a n a g a n J. L. 1960. Models for approximating basilar membrana displacement. Bell. Syst. Techn. J., 39 : 1163—1191.
- H u g g i n s W., J. L i c k l i d e r. 1951. Place mechanisms of auditory frequency analysis. J. Acoust. Soc. Amer., 23 : 290—299.
- K a t s u k i Y., T. S u m i, H. U c h i y a m a, T. W a t a n a b e. 1958. Electric responses of auditory neurons in cat to sound stimulation. J. Neurophysiol., 21 : 569—588.
- Z w i s l o c k i J. 1950. Theory of the acoustical action of the cochlea. J. Acoust. Soc. Amer., 22 : 778—784.

МОДЕЛЬ ДЛЯ ОПИСАНИЯ РЕЗУЛЬТАТОВ ПСИХОАКУСТИЧЕСКИХ ЭКСПЕРИМЕНТОВ СО СТАЦИОНАРНЫМИ СИГНАЛАМИ

В. Л. Темов

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

К настоящему времени накоплено очень много экспериментальных данных, относящихся к восприятию человеком простых акустических сигналов, но не существует теории, которая позволила бы количественно предсказывать результаты психоакустических экспериментов со сложными звуками. Создание достаточно полной «физиологической модели» прохождения сигналов в слуховой системе, модели, блоки которой соответствовали бы известным анатомическим структурам, является пока лишь благим пожеланием: недостаток детальной информации о многих про-

цессах приводит к тому, что любая модель, претендующая на сколько-нибудь полное объяснение экспериментальных данных, вынуждена опираться на допущения, недоступные опытной проверке. Сложность процессов приводит к громоздким моделям, которые допускают лишь качественные выводы, так как их реализация в виде физического аналога или программы для ЭВМ пока практически неосуществима. Эти обстоятельства оправдывают попытку построения чисто феноменологической модели психоакустических явлений, основными требованиями к которой были бы реальная возможность проведения расчетов и приемлемая точность совпадения их результатов с экспериментом. Особенно важно, чтобы модель позволяла «вычислить» восприятие человеком сложных стимулов, так как без этого невозможно применять данные психоакустики к исследованию восприятия речи.

За основу следует взять именно психоакустические эксперименты, так как в них вход полностью контролируется, а соответствующий ему выход (на уровне психологического решения) может быть измерен объективно и с высокой точностью. Психоакустические закономерности доступны для моделирования, и в то же время многие из них отражают фундаментальные особенности работы слухового механизма.

В настоящей работе описывается попытка построения модели обработки стационарных сигналов (т. е. достаточно длительных сигналов с постоянным во времени спектром) и приводятся результаты расчетов, выполненных для этой модели.

При разработке модели предполагалось, что анализ стационарного сигнала со сложным спектром можно разбить на два этапа: частотный анализ (в модели он осуществляется при помощи гребенки фильтров) и процедуру выделения некоторых особенностей спектра, таких, как максимумы, края полосовых шумов и другие существенные неоднородности. Допущение, что слуховое описание сложного спектра сводится к указанию частот и относительных интенсивностей такого рода особенностей, распространено среди исследователей речи и нашло свое отражение в создании формантных вокодеров.

Предполагалось также, что ряд психоакустических феноменов, связываемых с так называемыми критическими полосами, и феномен одновременной маскировки являются автоматическим следствием именно такого механизма обработки сигнала, а получаемые в экспериментах количественные зависимости отражают параметры его блоков.

И в самом деле, естественно ожидать, что следствием конечной ширины фильтров входной гребенки будет изменение порога восприятия одного стимула в присутствии другого по сравнению с его восприятием в тишине, т. е. маскировка первого стимула вторым. С другой стороны, максимум огибающей спектра является локальной его характеристикой, так что механизм выделения максимума должен обрабатывать спектр на сравнительно малом частотном интервале, и работа такого механизма должна приводить к ряду эффектов, связанных с тем, что в узкой области частот акустическая информация обрабатывается существенно по-иному, чем в широкой. Такие эффекты естественно сопоставить с теми, которые в психоакустике связывают с критическими полосами (Фрейдин, 1968). С шириной критической полосы обычно

связывают предельное расстояние по частоте между двумя составляющими сложного сигнала, которые различаются в психоакустических экспериментах тем или иным способом (Plomp, 1964). Если предположить, что воспринимаемым как различные составляющим соответствует выделение двух максимумов в спектре, а сливающимся — одного, то можно сделать вывод, что механизм критической полосы существенно связан с механизмом выделения максимумов. В силу этого в рассматриваемой модели эти механизмы просто отождествлены.

Таким образом, модель должна воспроизводить кривые одновременной взаимной маскировки (все сигналы — стационарные) и выделять максимумы в спектре при помощи механизма особой обработки информации в пределах относительно узких «критических» полос (по частоте). Выбрав структурную схему модели таким образом, чтобы она удовлетворяла этим условиям, необходимо подобрать ее параметры для достижения количественного совпадения с результатами экспериментов. При разработке структурной схемы учитывались качественные зависимости, обнаруженные во многих весьма разнообразных психоакустических исследованиях, однако в подгонке параметров учитывались лишь некоторые эксперименты, в то время как остальные считались контрольными.

ОБЩЕЕ ОПИСАНИЕ МОДЕЛИ

Модель реализуется в виде программы для ЭВМ, что предопределяет ее дискретный характер. Общая схема модели изображена в левой части рис. 1. Информация в модели преобразуется последовательно отдельными одномерными слоями. Каждый слой

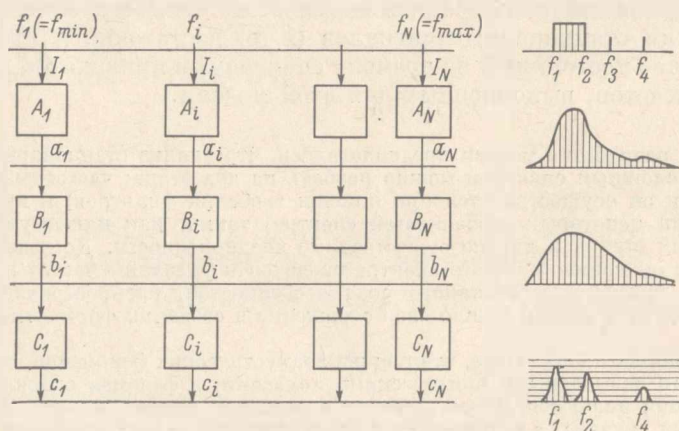


Рис. 1. Общая схема модели и картинки на выходах слоев.

Объяснение в тексте.

построен из N однотипных элементов, параметры которых, вообще говоря, могут меняться от первого элемента к N -му. Каждый элемент слоя в соответствии с «принципом места» отвечает определенной характеристической частоте f_i . Частоты f_i образуют возрастающую последовательность, покрывающую частотный диапазон от $f_1 = f_{\min}$ до $f_N = f_{\max}$. Совокупность выходных величин

элементов слоя, перечисленных в порядке возрастания характеристической частоты, будем называть картинкой на выходе данного слоя (или на входе следующего). Таким образом, каждый слой преобразует входную картинку в выходную. Входом всей системы является набор мощностей I_i спектральных составляющих (стационарного) входного сигнала. Элементы A_i , образующие первый слой, в совокупности образуют гребенку линейных фильтров. Выход элемента A_i выражается через значения многих входных гармоник I_j по формуле

$$a_i = \sum_{j=1}^N a_{ij} I_j. \quad (1)$$

Параметры a_{ij} определяются частотной характеристикой фильтра A_i . Элементы второго слоя B_i осуществляют логарифмическое преобразование. Выходной сигнал элемента B_i зависит только от одного a_i :

$$b_i = \begin{cases} \log b_i / \gamma_i, & \text{если } b_i \geq \gamma_i, \\ 0, & \text{если } b_i < \gamma_i, \end{cases} \quad (2)$$

где γ_i — аналог абсолютного порога восприятия тона на частоте f_i . В значения параметров входной отчасти информация о частотной характеристике входного микрофона (т. е. тракта от входа до гребенки фильтров), так как порог восприятия зависит и от нее.

Выход каждого из элементов C_i , образующих третий слой, является функцией некоторого количества b_j , расположенных вблизи C_i , т. е. таких, что $|j-i| \ll N$. Иначе говоря, выход C_i определяется лишь частью входной картинки слоя, расположенной недалеко от этого элемента. Значение выходного сигнала определяется следующим образом. Обозначим

$$d_i = \sum_j \delta_{ij} b_j, \quad (3)$$

где δ_{ij} — параметры элемента C_i , отличные от нуля только для $|j-i| \ll N$. Тогда

$$c_i = \begin{cases} d_i - p, & \text{если } d_i \geq p, \\ 0, & \text{если } d_i < p. \end{cases} \quad (4)$$

Здесь p — параметр, имеющий смысл порога срабатывания, общий для всех элементов слоя C . Параметры этих элементов подобраны таким образом, что на выходной картинке слоя подчеркиваются максимумы в спектре входного сигнала и края полосовых шумов.

На рис. 1 справа схематично изображено прохождение через модель сигнала, состоящего из шума в полосе частот от f_1 до f_2 и двух тонов с частотами f_3 и f_4 . В результате на выходе модель

выделяет три «максимума» — на частотах f_1 , f_2 и f_4 , соответствующих срезам полосового шума и слышимой на его фоне гармонике. Гармоника частоты f_3 в данном случае «замаскирована».

Для сравнения работы модели с результатами экспериментов необходимо условиться относительно способа интерпретации выходного сигнала модели, т. е. процесса принятия решения. Основным, очевидно, является решение о тождественности либо различии двух сигналов.

Выходную картинку (т. е. совокупность величин c_i) мы будем рассматривать как сокращенную запись входного сигнала. Решение о различии или сходстве сигналов будет приниматься по выходной картинке. Возможны различные подходы к построению критерия различия, но здесь мы приведем лишь тот критерий, который использовался в экспериментах, описанных в конце настоящей работы.

Будем считать, что два сигнала S_1 и S_2 различаются моделью, если хотя бы для одного i будет

$$|c_i(S_1) - c_i(S_2)| \geq r,$$

и не различаются, если для всех i

$$|c_i(S_1) - c_i(S_2)| < r, \quad (5)$$

где r — некоторая пороговая величина.

Используя этот (или другой сходный) критерий различимости сигналов и задавшись некоторой величиной ϵ , характеризующей точность, мы можем для двух произвольных сигналов M и S найти такое $\lambda > 0$, что сигналы M и $M + \lambda S$ не различаются моделью, а M и $M + (\lambda + \epsilon) S$ — различаются. Очевидно, это эквивалентно тому, чтобы найти уровень маскировки сигнала S маскером M .

Выбрав критерий различия, можно построить кривые маскировки, например тона тоном или тона шумом, или любых других сигналов в зависимости от тех или иных условий.

ПАРАМЕТРЫ МОДЕЛИ

Для однозначного описания модели необходимо определить ее параметры f_i , α_{ij} , γ_i , δ_{ij} , p , r . При выборе параметров естественно стремиться к тому, чтобы их было как можно меньше или, точнее, чтобы их можно было выразить через минимальное количество независимых параметров. Другое естественное стремление — провести оценку параметров по результатам минимального числа экспериментов, сделав остальные эксперименты контрольными.

Параметры будем оценивать, предполагая, что структурная схема модели позволяет в принципе с достаточной точностью описывать результаты психоакустических экспериментов со ста-

ционными сигналами. Забегая вперед, можно констатировать, что это предположение в основном подтвердилось.

Прежде всего необходимо выбрать закон возрастания последовательности $f_1, \dots, f_i, \dots, f_N$. Лучший выбор, очевидно, должен быть таким, чтобы обеспечить одинаковую форму как для всех фильтров A_i , так и для всех элементов c_i , если только это возможно. В этом случае коэффициенты α_{ij} и δ_{ij} должны зависеть лишь от разности индексов

$$\alpha_{ij} = \alpha(j - i), \quad (6a)$$

$$\delta_{ij} = \delta(j - i). \quad (6b)$$

Обе зависимости (6) могут, оказывается, выполняться одновременно при условии, что последовательность f_i меняется равномерно по шкале тональностей z , или шкале барков (Zwicker, 1961). В самом деле, характеристики фильтров A_i определяют в существенных чертах ход кривых маскировки тона тоном при изменении частот тона и маскера. (Это достаточно очевидно и легко проверяется непосредственным расчетом). Ввиду этого сами характеристики фильтров в модели выбираются исходя именно из кривых маскировки тона тоном. Но эти последние, изображенные в шкале тональностей, выглядят почти тождественно для всех частот маскера (Maiwald, 1967), т. е. зависят лишь от разности тональностей маскера и маскируемого тона. С другой стороны, элементы C_i соотносятся в модели с механизмом критических полос (в дальнейшем это будет обсуждено подробнее). Но шкала тональностей по своему построению такова, что единицей расстояния по ней является критическая полоса. В результате условие (6b) выполняется естественным образом.

Важность условий (6a) и (6b) состоит в том, что они уменьшают количество независимых параметров α_{ij} и δ_{ij} ровно в N раз, так как матрица $\{\alpha_{ij}\}$ (и аналогично $\{\delta_{ij}\}$) теперь может быть однозначно построена, если задать всего лишь одну ее строку.

При составлении программы для ЭВМ удобно иметь выражение тональности в барках через частоту и обратное выражение в виде несложной аналитической зависимости. На рис. 2 показана кривая $B = B(f)$ и график аппроксимирующей зависимости

$$B = 6.7 \operatorname{Arsh} \frac{f - 20}{600}. \quad (7)$$

Расхождение между аппроксимируемым и аппроксимирующим выражениями в диапазоне от 150 до 5000 гц не превосходит 3% по тональности и 5% по частоте. Исходя из (7), можно вывести следующие соотношения между номером точки, ее характеристической частотой и тональностью:

$$f_i = 600 \operatorname{sh} (B_i / 6.7) + 20;$$

$$B_i = B_{\min} + \frac{i-1}{N-1} (B_{\max} - B_{\min}) = 6.7 \left[\operatorname{Arsh} \frac{f_{\min} - 20}{600} + \right. \\ \left. + \frac{i-1}{N-1} \left(\operatorname{Arsh} \frac{f_{\max} - 20}{600} - \operatorname{Arsh} \frac{f_{\min} - 20}{600} \right) \right];$$

$$f_i = 600 \operatorname{sh} \left[\operatorname{Arsh} \frac{f_{\min} - 20}{600} + \frac{i-1}{N-1} \left(\operatorname{Arsh} \frac{f_{\max} - 20}{600} - \right. \right. \\ \left. \left. - \operatorname{Arsh} \frac{f_{\min} - 20}{600} \right) \right] + 20;$$

$$i = 1 + (N-1) \frac{B_i - B_{\min}}{B_{\max} - B_{\min}} =$$

$$= 1 + (N-1) \frac{\operatorname{Arsh} (f_i - 20)/600 - \operatorname{Arsh} (f_{\min} - 20)/600}{\operatorname{Arsh} (f_{\max} - 20)/600 - \operatorname{Arsh} (f_{\min} - 20)/600}.$$

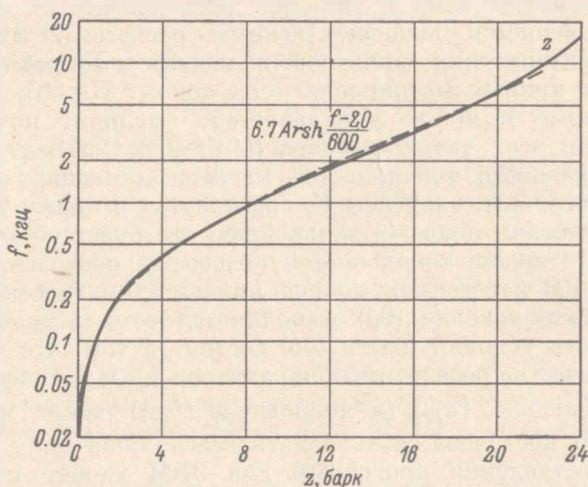


Рис. 2. Сравнение зависимости $z=B(f)$ и аппроксимирующего выражения.

По оси абсцисс — тональность z , барк; по оси ординат — частота f , кгц. Сплошная линия — зависимость $z(f)$ по Цвиккеру (Zwicker, 1961), прерывистая — аппроксимирующее выражение (7).

Зависимость между частотой и тональностью приблизительно линейная для низких частот ($f_i < 500$ гц) и приблизительно логарифмическая для высоких ($f_i > 1200$ гц).

Перейдем теперь к гребенке фильтров A_i . Как уже упоминалось, частотная характеристика A_i определяет зависимость

уровня маскировки тона тоном частоты f_s . Действительно, пусть значение a_j для некоторого (маскирующего) тона частоты f_j заметно больше значения a_j для чистого тона частоты f_j . Тогда при одновременной подаче на вход обеих гармоник на выходной картинке слоя A (и тем более слоя B) не будет никаких заметных особенностей на частоте f_j . Если же соотношение значений a_j для тонов будет обратным, то на частоте f_j будет заметный максимум. Следовательно, пороговое значение, при котором новый максимум на частоте f_j будет едва-едва выделяться на выходе слоя C и обнаруживаться, по критерию (5), для данного уровня маскера целиком определяется характеристикой фильтра, т. е. a_{j_0} . Такой вывод, очевидно, справедлив независимо от способа обнаружения максимума слоем C , лишь бы этот способ обнаруживал достаточно большие максимумы и не реагировал на очень маленькие, т. е. обладал бы наряду с чувствительностью некоторой помехоустойчивостью. Зависимости маскировки тона тоном для различных частот хорошо исследованы (Чистович, 1957; Small, 1959; Шупляков, 1967; Maiwald, 1967). Общая закономерность состоит в том, что для $f_s < f_m$ маскировка заметна лишь на определенном интервале от f_m по оси частот. При увеличении интенсивности маскера этот интервал медленно растет, оставаясь меньше ширины критической полосы. В терминах модели это значит, что для $f_s < f_m$ маскировка определяется в основном элементами слоя C , действие каждого из которых распространяется на ограниченный интервал, а фильтры A_i должны иметь достаточно крутой спад в сторону высоких частот. Можно принять крутизну спада близкой к максимальной крутизне кривых маскировки, наблюдавшейся в экспериментах, т. е. 150—200 дБ на октаву. Забегая вперед, укажем, что лучше всего выходные кривые модели совпадают с экспериментальными при наклоне 30—35 дБ/барк, т. е. приблизительно 120—150 дБ на октаву на частотах выше 800 гц.

Для $f_s > f_m$ маскировка также нелинейна, но в значительно меньшей степени. Если пренебречь нелинейностью, крутизна спада частотной характеристики фильтров A_i в сторону низких частот должна равняться наклону кривых маскировки при средних интенсивностях маскера, т. е. 9—12 дБ/барк (приблизительно 30—40 дБ на октаву для высоких частот). Параметры γ_i элементов второго слоя B можно определить исходя из величины порога восприятия тонов в тишине. В качестве аппроксимации соответствующей кривой выбрано выражение

$$\gamma(B) = 40 \exp(-B/2.7) + 7 \exp((B-20)/4),$$

где B — тональность в барках, связанная с частотой соотношением (7). В диапазоне частот 50—5000 гц приведенное выра-

жение дает значения порога, хорошо совпадающие с экспериментальными данными.

Перейдем к параметрам элементов слоя C . Учитывая (6б), достаточно задать зависимость $\delta(j-i)$ для любого i . Вид функции $\delta(j-i)$, принятый в модели, изображен на рис. 3.

Основания для выбора такой зависимости, похожей на часто используемую в моделях латерального торможения, следующие. Непосредственным расчетом можно установить, что модель при этом действительно выделяет спектральные максимумы и края полосовых шумов. Маскировка тона более высокочастотным тоном оказывается заметной лишь в ограниченной области частот ниже частоты маскера (менее Δ барк — рис. 3). Наконец, такой же

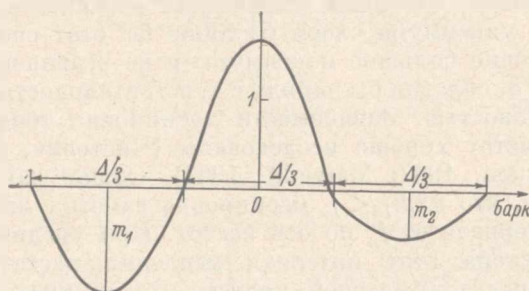


Рис. 3. Весовая функция элемента слоя C .

По оси абсцисс — тональность, барк (нуль соответствует характеристической частоте элемента); по оси ординат — весовые коэффициенты, отн. ед.

примерно вид имеет функция влияния тонального стимула на импульсацию в волокне слухового нерва кошки в присутствии тона характеристической для данного волокна частоты (Sachs, Kiang, 1968). Эти эксперименты дают основание и для выбора ширины функции Δ — она должна быть порядка 3 барков. Асимметрия, определяемая отношением параметров m_1 и m_2 , также может быть оценена по этим экспериментам.

Примем нормировку

$$\sum_{\{j: \delta_{ij} > 0\}} \delta_{ij} = 1$$

и рассмотрим, какова должна быть сумма всех весовых коэффициентов δ_{ij} . Пусть, например,

$$\sum_j \delta_{ij} = 0. \quad (8)$$

В этом случае маскировка тоном более высокочастотных тонов будет линейной. Действительно, при изменении входного сигнала на x дБ величины a_i изменяется в силу (1) в одно и то же число раз,

а величины b_i — в силу (2) на одно и то же количество y . Но в этом случае, согласно (3) и (4),

$$d_i(x) = \sum_j \delta_{ij} (b_j + y) = d_i + y \sum_j \delta_{ij}, \quad (9)$$

т. е. при нормировке (8) модель не чувствует изменений уровня (это относится к достаточно большим сигналам, так как при малых начинают играть роль пороги γ_i). Не изменится при этом и разность уровней маскера и сигнала на пороге обнаружения. Однако экспериментальные данные (Чистович, 1957; Small, 1959; Шупляков, 1967; Maiwald, 1967) говорят о том, что при увеличении уровня маскера разность уровней маскера и порогового сигнала уменьшается. Это уменьшение составляет в среднем 4—5 дБ на 20 дБ прироста маскера. Такое явление наблюдается, как видно из (9), если вместо (8) в модели принять

$$\sum_j \delta_{ij} = \alpha < 0. \quad (10)$$

В этом случае некоторые из выходов c_i , бывшие больше нуля для исходных значений маскера и порогового сигнала, могут стать равными нулю [см. (4)] при одновременном увеличении уровней маскера и сигнала. Это эквивалентно, очевидно, увеличению маскировки, так как чтобы обнаружить сигнал в новых условиях, необходимо увеличить его уровень на большую величину, чем увеличился маскер. Подбор α в (10) лучше всего произвести в процессе экспериментальных расчетов, так же как и подбор пороговых параметров p и r в (4) и (5).

ЭКСПЕРИМЕНТЫ

Описанная модель была реализована в виде программы для ЭВМ. Модель позволяла работать со стационарными сигналами, заданными набором мощностей спектральных составляющих. Обычно расчеты проводились для $N=200$ и $N=300$ точек. На выходе получался набор выделенных особенностей спектра. Для любых двух сигналов можно было определить маскировки одного из них другим, а также уровень одного из сигналов, необходимый, чтобы замаскировать другой. Ниже описаны результаты трех экспериментов с моделью, поставленных аналогично известным психоакустическим экспериментам.

В первом эксперименте определялась маскировка тонов тоном частоты 1000 гц. Полученные результаты приведены на рис. 4. Видно, что в целом соответствие вполне удовлетворительное. Особенно интересно отметить появление горбов на кривых справа от маскера. Эти горбы отстоят от маскера не на октаву, а гораздо меньше — приблизительно на 0.7 октавы, что не позволяет отнести их за счет появления гармоник основного тона, субъек-

тивных либо вызванных нелинейностью механической части слухового тракта. И в эксперименте, и в модели они появляются при больших интенсивностях маскера. Уровень маскера, при котором начинают появляться горбы на кривых маскировки, полученных на модели, можно легко изменять, задавая α в выражении (10). На рис. 4 намек на горб есть уже для уровня маскера $I_m = 60$ дБ (в этом расчете было принято $\alpha = 0.11$), а при $\alpha = 0.04$ горб появляется лишь при $I_m = 100$ дБ. Следует отметить, однако,

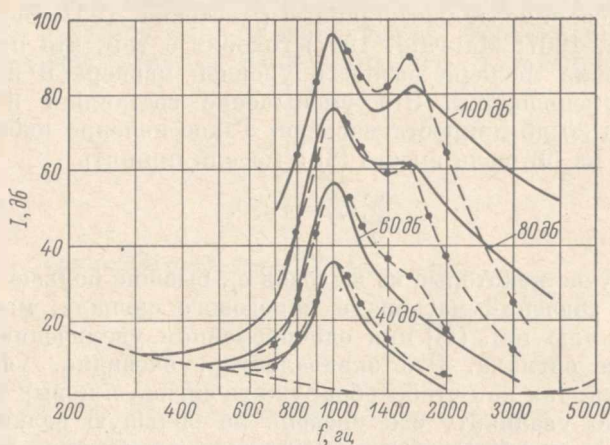


Рис. 4. Маскировка тона 1000 гц — сравнение данных, полученных на модели и в психоакустическом эксперименте.

По оси абсцисс — частота маскируемого тона, гц; по оси ординат — интенсивность тона, дБ от $2 \cdot 10^{-4}$ бар/см². Сплошные кривые — данные психоакустического эксперимента (Maiwald, 1967), прерывистые — расчетные значения тона на пороге обнаружения, нижняя прерывистая кривая — порог обнаружения в тишине. Параметр кривых — интенсивность маскера.

что правее горба экспериментальные кривые спадают гораздо круче, чем расчетные. Очевидно, здесь участвует какой-то механизм, не отраженный в модели. Совпадение левой части кривых вполне удовлетворительное. Психоакустическим аналогом второго эксперимента были эксперименты по определению различимости составляющих сложного звука (Plomp, 1964). В этой работе, в частности, получилось, что при расстоянии между гармониками, равном 100 гц, их выделение в сложном сигнале возможно лишь в случае, если сами частоты гармоник не превышают 500—600 гц. На модели вычислялась выходная картинка, соответствующая сигналу, состоящему из 20 равноинтенсивных (60 дБ) гармоник тона 100 гц. Эта картинка изображена на рис. 5. Выделенные моделью особенности спектра соответствуют частотам 100, 200,

300, 400, 500 и 2000 гц. Первые 5 из них следует интерпретировать как различные гармоники, последнюю — как край полосового шума. При проведении расчета значение Δ пришлось уменьшить до 2.5 барка по сравнению с 3.5 барка в предшествующем случае. Такое изменение согласуется с результатами обсуждаемой работы, где предельное расстояние между различаемыми гармониками получилось равным критической полосе для частоты выше 1000 гц и равным около 0.7 критической полосы для частот ниже 500 гц. Положение можно исправить, если перейти от шкалы тональностей к другой шкале, в которой низкие частоты будут соответственно растянуты.

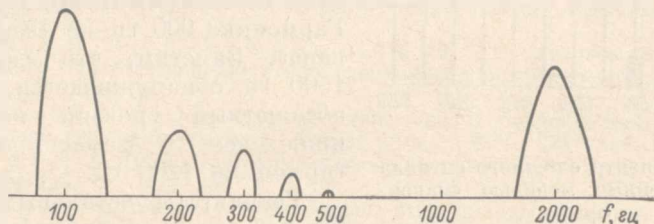


Рис. 5. Выход модели, соответствующий входному сигналу, состоящему из 20 равноинтенсивных гармоник тона 100 гц.

По оси абсцисс — частота, гц; по оси ординат — выход модели, отн. ед.

Психоакустическим аналогом третьего эксперимента были эксперименты по восприятию синтетических гласных, спектры которых состоят из 10 гармоник тона 300 гц (Мущников, Чистович, 1971). В работе измерялся порог изменения фонетического качества звука при изменении одной из гармоник для двух типов спектров. Ввиду того что эта работа помещена в настоящем сборнике, нет необходимости останавливаться здесь на методике эксперимента. Отметим лишь, что результаты можно интерпретировать как аргумент, во-первых, за то, что восприятие использованных стимулов в основном определяется взаимной маскировкой составляющих и, во-вторых, что изменению фонемного качества соответствует появление новой особенности в спектре (обычно нового максимума, но иногда и просто заметной неоднородности). Интерпретируя эти данные в терминах модели, следует ожидать, что изменению фонемного качества сигнала должно сопутствовать появление на выходной картинке нового максимума на частоте, соответствующей изменяемой гармонике. Такое предположение полностью подтвердилось. Из десяти просчитанных экспериментов в восьми, где изменяемая гармоника имела частоты от 900 до 1800 гц, дополнительный максимум появлялся в соответствующем месте. В двух случаях, когда изменяемые гармоники имели частоты 2100 и 2400 гц, для появления максимума не хватало соот-

ветственно 0.6 и 2 дб (при интенсивностях гармоник 47.2 и 44.6 дб).

Следует отметить, что результаты этого расчета оказались мало критичными к изменению большей части параметров модели. На рис. 6 изображен спектр одного из двух типов сигналов,

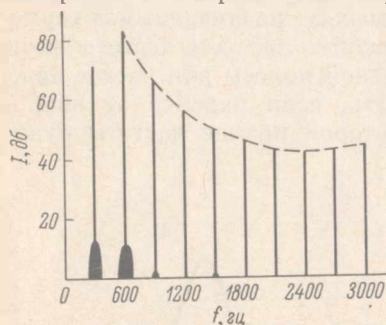


Рис. 6. Спектр сложного сигнала и выделенные моделью «максимумы».

По оси абсцисс — частота, гц; по оси ординат — интенсивность гармоник, дб. Вертикальные линии — интенсивности гармоник (прерывистая линия по частоте 1500 гц — уровень, соответствующий изменению фоновому качества); зачерненные столбики внизу — выход модели.

использовавшихся в упомянутой работе. Уровень гармоник 1500 гц, при котором она начинает обнаруживаться человеком по изменению фоновому качества, показан прерывистой линией. Внизу показаны максимумы на выходе слоя С. Гармоника 900 гц не «замаскирована». Заметим, что гармоника 1500 гц обнаруживается, хотя ее абсолютный уровень несколько ниже, чем у «замаскированной» гармоник 1200 гц.

Результаты остальных 9 опытов выглядят аналогично. Одной из причин указанного выше несовпадения в двух случаях (из десяти) может быть тот факт, что параметры модели отражают характеристики «среднелитературного» испытуемого, которые мо-

гут отличаться от характеристик испытуемых, принимавших участие в обсуждаемых экспериментах.

Здесь уместно отметить, что при характеристиках фильтров слоя А, соответствующих характеристикам так называемых эквивалентных фильтров улитки, как показано в работе Лабутина и Павловского (1971), этот же набор сигналов практически невозможно различить никакой последующей обработкой.

ОБСУЖДЕНИЕ

Подбор параметров модели, описанной в настоящей работе, производился путем подгонки к результатам экспериментов по маскировке тона тоном. Этого оказалось достаточно, чтобы модель демонстрировала вполне удовлетворительную работу на довольно сложных сигналах, состоящих из большого числа (10—20) гармоник.

Можно считать, что полученные результаты говорят в пользу предположения о том, что описание сложного сигнала с помощью перечисления положения и интенсивности особенностей его спектра (типа максимумов) является достаточным для их идентификации. При этом следует указывать только максимумы, не замаскированные другими составляющими сигнала.

Модели подобного типа, после необходимых уточнений и обобщений, можно использовать для вычисления результатов психоакустических экспериментов. Особенно важной представляется возможность работать со сложными сигналами. Создание таких моделей позволяет обобщать психоакустические данные, собранные при исследовании простых стимулов, на случай сложных стимулов, в том числе некоторых речевых.

Л и т е р а т у р а

- Л а б у т и н В. К., В. В. П а в л о в с к и й. 1971. Изучение реакции основной мембраны улитки на сложные акустические стимулы. Наст. сб. : 29—36.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971. О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном. Наст. сб. : 11—19.
- Ф р е й д и н А. А. 1968. О роли критических полос в обработке информации слуховой системой человека. Акуст. журн., XIV : 321—336.
- Ч и с т о в и ч Л. А. 1957. Частотные характеристики эффекта маскировки. Биофизика, 2 : 743—755.
- Ш у п л я к о в В. С. 1967. Слуховой анализ стационарных шумных согласных. Дисс. Л.
- M a i w a l d D. 1967. Beziehungen zwischen Schallspektrum Mithörschwelle und der Erregung des Gehörs. Acustica, 18 : 69—80.
- P l o m p R. 1964. The ear as a frequency analyser. J. Acoust. Soc. Amer., 36 : 1628—1636.
- S a c h s M. B., N. Y. S. K i a n g. 1968. Two-tone inhibition in auditory-nerve fibers. J. Acoust. Soc. Amer., 43 : 1120—1128.
- S m a l l A. M. 1959. Pure-tone masking. J. Acoust. Soc. Amer., 31 : 1619—1625.
- Z w i c k e r E. 1961. Subdivision of the audible frequency range into critical bands. J. Acoust. Soc. Amer., 33 : 248.

ВОСПРИЯТИЕ КОРОТКИХ ОТРЕЗКОВ ГЛАСНЫХ ЗВУКОВ

Б. В. Богданов

Институт эволюционной физиологии и биохимии им. И. М. Сеченова АН СССР, Ленинград

Распространено предположение, что информацией, используемой при распознавании гласного в речевом потоке, являются те или иные характеристики спектра (отношение энергии в полосах, значения частот формант) на стационарном участке или в точках «цели» (target value).

Так как длительность стационарного участка — положения близкого к «цели» — может быть очень короткой, эта гипотеза предполагает, что слуховая обработка спектра должна быть мало-

инерционной. Психоакустические данные (Brady et al., 1961; Heinz et al., 1967) соответствуют представлению о малой инерционности слуховой обработки.

В противоречии с этим находится отмечавшийся рядом исследователей (Ahmend, Fatehchand, 1959) факт, что укорочение гласных до 60—40 мсек. приводит к резкому снижению их опознаваемости.

В настоящей работе была сделана попытка проверить предположение, что это снижение разборчивости связано не с инерционностью слуховой обработки спектров, но с отсутствием ассоциаций между необычно короткими по длительности сигналами и гласными фонемами.

МЕТОДИКА

Мы располагали списком гласных русской речи в 30 звукосокращениях типа СГС, составленным фонетистами ЛГУ.

<i>фул</i>	<i>кор</i>	<i>фат</i>	<i>бэк</i>	<i>вый</i>
<i>гун</i>	<i>сот</i>	<i>сай</i>	<i>пэм</i>	<i>тыщ</i>
<i>нус</i>	<i>лой</i>	<i>лан</i>	<i>тэл</i>	<i>мыр</i>
<i>муф</i>	<i>моц</i>	<i>маш</i>	<i>шэс</i>	<i>мыш</i>
<i>шуш</i>	<i>зох</i>	<i>пак</i>	<i>шэт</i>	<i>шым</i>
<i>лут</i>	<i>жон</i>	<i>цар</i>	<i>пэр</i>	<i>цын</i>

Как видно из приведенного списка звукосокращений, каждый из 5 гласных звуков встречается 6 раз и находится между твердыми согласными. Гласный [i], не удовлетворяющий последнему условию, не исследовался. Список был прочитан перед конденсаторным микрофоном двумя дикторами (мужчина и женщина), записан на магнитофонную пленку и размножен в копиях. Магнитофонная пленка разрезалась соответственно границам звукосокращения и склеивалась в кольца. Перед звукосокращением в каждом кольце записывался короткий высокочастотный (12 кгц) импульс, который служил в качестве управляющего сигнала пусковому устройству сепаратора — прибора для выделения звуков из слов и звукосокращений. С помощью этого прибора из той части гласного, которая условно называется квазистационарной из-за относительно устойчивых акустических характеристик, видимых на спектрограмме, выделялись звуковые отрезки длительностью в 10, 15, 20, 30, 40, 60, 80, 100, 120, 140 мсек. Срединная часть квазистационарного участка гласного¹ использовалась для выделения коротких звуковых отрезков длительностью в 10, 15, 20, 30, 40 и 60 мсек. Считалось, что такой способ получения их является в какой-то мере гарантией того, что звучание коротких отрезков будет достаточно «чистое», так как они находятся вне области случайного влияния переходного процесса в гласном. Каждому из 10 упомянутых выше значений длительности соответствовала отдельная запись на магнитофонной ленте серии, состоящей из 30 звуков. Паузы между звуками были в пределах от 3 до 3.5 сек. Таким образом, в общей сложности в качестве исходного материала для прослушивания было получено 20 основных серий записи (по 10 серий от каждого диктора). Некоторые из этих серий имели по 3 и 4 варианта записи, которые отличались от основной серии лишь порядком следования звуков. Как в основных сериях, так и в вариантах порядок следования звуков был случайным.

¹ Средняя величина длительности квазистационарного участка гласных равна 130 мсек. (по двум дикторам).

Одновременно с составлением серий записи по отдельным длительностям решались вопросы, связанные с унификацией исходного материала исследования. Особенно это касалось коротких звуков длительностью в 10, 15, 20, 30, 40 и 60 мсек. Во-первых, существенно необходимым условием опытов являлось условие равного уровня громкости звуков внутри серии записи. Во-вторых, для исключения возможного влияния на восприятие начальной фазы короткого звука во всех конкретных случаях был найден нуль фазы основной частоты квазистационарного участка гласного, от которого и отсчитывалась искомая величина длительности. Как первое, так и второе решалось путем сопоставления слухового и осциллографического анализов.

В опытах по прослушиванию участвовало 13 аудиторов (мужчин и женщин) с нормальным слухом. Прослушивание производилось бинаурально с помощью телефонов типа ТД-6.

Помимо магнитофонов МЭЗ-28 и сепаратора, в опытах использовались также динамический спектрограф, универсальный чернильный самописец типа УСЧ-8-02, конденсаторный микрофон НТБ № 5 и ламповые вольтметры.

РЕЗУЛЬТАТЫ ОПЫТОВ

Опыты по прослушиванию 10 основных серий записи начались с тренировки аудиторов. Инструкцией служила следующая фраза экспериментатора: «Сейчас вы услышите 30 звуковых последовательностей, которые соответствуют русским гласным; запишите их на бланке буквами, независимо от Вашей уверенности в правильности принятого решения». После опыта бланк с ответами отбирался, данные по опознаваемости не сообщались аудитору. Нетрудно видеть, что в инструкции не было сведений о количественном составе гласных, частоте их встречаемости и длительности.

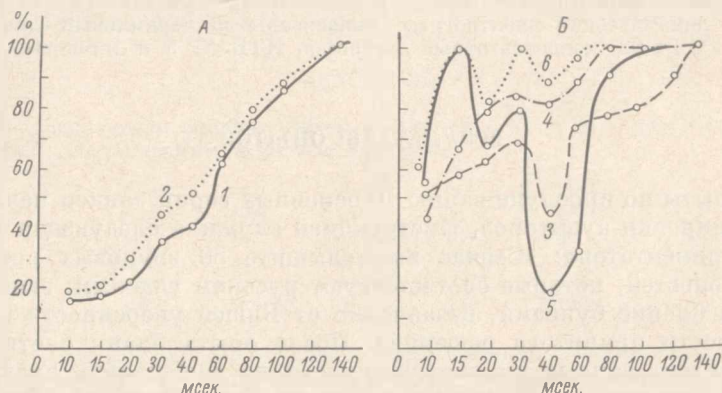
Характер инструкции и то, что результаты по опознаваемости не сообщались аудитору, отвечали общему замыслу эксперимента; опыты должны были проходить в условиях самообучения. Иначе говоря, опыты должны были проводиться таким образом, чтобы, с одной стороны, исключить трудно учитываемые влияния инструкции на результаты по опознаваемости и, с другой стороны, максимально использовать (при анализе результатов) индивидуальные способности аудиторов к обучению на основе их собственного жизненного опыта.

Первые 5 тренировок по прослушиванию были вводными. Результаты последующих 5—7 и 30—40 тренировок по опознаванию 10 серий записи усреднены и отражены на рисунке (А).

На этом этапе работы данные по опознаваемости ничем не отличались от данных, полученных ранее другими исследователями. Однако после 150—250 тренировок характер хода кривой по опознаваемости существенно изменился. Кривая 3 (см. рисунок, Б) отражает эти изменения. В ходе кривой 3 по сравнению с ходом предыдущих кривых резко обозначилось уменьшение опознаваемости (более чем на 20%) звуков длительностью в 40 мсек., которое, однако, в своем суммарном значении не падало

ниже уровня, достигнутого на первых этапах тренировок. Изменения в опознаваемости в ходе тренировок требовали объяснения.

Прежде всего в контрольном опыте была проверена опознаваемость этой серии записи на аудиторе, который не принимал участия в предыдущих опытах. После 10—12 тренировок уровень опознаваемости 40-мсек. звука достиг 60%. Опыт этот показал, что для понимания изменений в опознаваемости коротких звуков у 13 аудиторов необходим анализ в деталях результатов их опознаваемости, особенностей условий и хода экспериментов по прослушиванию.



Опознаваемость гласных в зависимости от длительности их квазистационарного участка.

По оси абсцисс — длительность отрезков гласных, мсек.; по оси ординат — опознаваемость гласных звуков, %. А — результаты, полученные при числе тренировок менее 40; Б — при числе тренировок 150 и более. 1 — суммарные данные по ответам 13 аудиторов после 5—7 тренировок; 2 — по ответам 13 аудиторов после 30—40 тренировок; 3 — по ответам 13 аудиторов после 150—250 тренировок; 4 — по ответам 5 аудиторов после 250—300 тренировок; 5 — по ответам 2 аудиторов после 350—370 тренировок; 6 — по ответам 2 аудиторов после 357—377 тренировок.

В итоге этой работы было выяснено, что не все аудиторы в равной мере способны успешно продолжать тренировки по опознаванию коротких звуков. По этому критерию 13 аудиторов были разделены на две группы. В первую группу входили 8 аудиторов (с числом тренировок от 250 до 270), которые слабо обнаруживали свои способности в опознавании коротких звуков (от 10 до 60 мсек. включительно). Во вторую группу входило 5 аудиторов (с числом тренировок от 120 до 130), которые более успешно справлялись с задачей опытов.

Динамика процесса изменения опознаваемости в ходе тренировок у аудиторов первой и второй групп существенно разная. Так, у аудиторов первой группы процесс изменения опознаваемости можно представить как постепенный и равномерный сдвиг вверх кривой 1 до уровня кривой 3. Этот сдвиг потребовал более

250 тренировок. У аудиторов второй группы тот же уровень был достигнут всего лишь за 120—130 тренировок, но с обязательным глубоким провалом в ходе кривой опознаваемости в области 40 мсек. Таким образом, вид кривой 3, суммарно отражающей опознаваемость обеих групп, обусловлен главным образом особенностями восприятия аудиторов второй группы.

После разделения аудиторов на две группы дальнейшая постановка опытов также претерпела некоторые изменения. Аудиторы первой группы, как и прежде, продолжали свои тренировки на всех 10 основных сериях записи коротких звуков, аудиторы второй группы продолжали свои тренировки только на одной серии записи — со звуками длительностью 15 мсек. Мотивом к такого рода изменениям в опытах со второй группой аудиторов послужило то обстоятельство, что прежняя постановка опытов не давала возможности судить о степени влияния смежных серий записей коротких звуков друг на друга, хотя априори факт такого взаимовлияния для нас был вне сомнений.

Выбор короткого звука длительностью в 15 мсек. в качестве объекта для прослушивания был сделан лишь потому, что в нем укладывалось два периода основной частоты голоса, поскольку последняя у дикторов была около 140—150 гц.

Последующая длительная (150—170 раз) тренировка аудиторов первой и второй групп привела к тому, что у аудиторов первой группы прирост опознаваемости всех 10 основных серий записи составлял единицы процентов от уровня кривой 3. Аудиторы второй группы, которые продолжали свои тренировки только на одной серии записи звуков длительностью 15 мсек., имели средний относительный прирост опознаваемости на 10—12%, т. е. в несколько раз больше, чем у аудиторов первой группы. Чтобы проверить у аудиторов второй группы влияние тренировки на 15-мсек. звуке на опознаваемость звуков других длительностей, был поставлен контрольный опыт по опознаванию всех 10 основных серий записей звуков разной длительности. Ход кривой 4 отражает полученные в этом опыте данные.

Относительно деталей хода кривой 4 по сравнению с ходом кривой 3 можно заметить следующее: во-первых, несколько уменьшилась опознаваемость звуков длительностью 10 мсек., зато опознаваемость звуков в 20 и 30 мсек. увеличилась значительно больше, чем опознаваемость звука в 15 мсек. (хотя тренировки на этих звуках с аудиторами второй группы не проводились); во-вторых, опознаваемость 80- и 100-мсек. звука увеличилась до 100%; в-третьих, вопреки ожиданиям, опознаваемость звука длительностью 40 мсек. увеличилась до 80%. В связи с последним фактом и тем, что характер хода кривых 3 и 4 сходен, была подробно изучена динамика изменений в опознаваемости у аудиторов этой группы. В итоге состав второй группы аудиторов вновь был пересмотрен. Из нее были исключены 3 аудитора по тому же

критерию, который был использован и раньше при делении 13 аудиторов на две группы, т. е. по величине и скорости прироста процента опознаваемости звуков длительностью 15 мсек. в зависимости от числа тренировок и наличия провала в ходе кривой опознаваемости в области 40 мсек.

На этом этапе опытов были образованы также две, но уже новые группы. Первая новая группа из 11 аудиторов (8+3) и вторая — из 2 аудиторов. Первая новая группа продолжала тренировки, как и прежде, на всех 10 основных сериях записей. Вторая новая группа продолжала тренировки лишь на одной серии записи звуков длительностью 15 мсек. После 100—120 тренировок оба аудитора этой группы достигли 100% уровня опознаваемости 15-мсек. звука, независимо от порядка предъявлений звуков и от того, был ли это голос диктора-мужчины или диктора-женщины.

Вновь в качестве контроля, применительно к этой группе аудиторов, был поставлен опыт по прослушиванию всех 10 основных серий записи звуков. Результаты его отражает кривая 5. Как видно из рисунка, основной результат опыта — увеличение опознаваемости звука 15-мсек. длительности до 100%. Как и прежде, в ходе кривой сохранился глубокий провал в области 40 мсек.

На последнем этапе исследований проводилась тренировка второй группы аудиторов по опознаванию звука длительностью 30 мсек. Интересно было выяснить, окажет ли она влияние на опознаваемость звуков длительностью 40 мсек. 100%-я опознаваемость звука длительностью 30 мсек. была достигнута всего лишь после 5—7 тренировок почти одновременно у обоих аудиторов. Затем им был предложен контрольный опыт по прослушиванию всех 10 основных серий записи звуков. Результаты последнего опыта отражены в виде кривой 6. Хотя точки с отметками 10, 20 и 40 мсек. не находятся на уровне 100% опознаваемости, однако для 100% опознаваемости их потребовалось всего лишь 5 или 6 дополнительных тренировок.

Таким образом, за исключением длительности 10 мсек., на остальных 9 основных сериях записей звуков двумя аудиторами из 13 была достигнута 100%-я опознаваемость.

ОБСУЖДЕНИЕ

В итоге проделанной работы были получены следующие фактические данные.

1. Короткие отрезки гласных звуков речи (длительностью в 15, 20, 30 мсек. и т. д.), полученные с помощью сепаратора из так называемых квазистационарных участков гласных звуков, идентифицируются человеком после длительной тренировки в 100% случаев как соответствующие гласные русской речи.

2. Длительные тренировки (18 месяцев) по опознаванию коротких гласных звуков речи показали, что существует два основ-

ных типа аудиторов: первый тип обнаруживает способность к 100%-й опознаваемости, второй тип не обнаруживает этой способности.

Чтобы оценить значимость результатов и средств, с помощью которых удалось их получить, следует обратиться к ряду фактов, добытых ранее другими исследователями.

До последнего времени твердо установилось мнение (Ahmend, Fatehchand, 1959), что для уверенной опознаваемости гласных звуков длительность отрезков их должна быть не менее 80—100 мсек. Вопрос о возможности опознавания отрезков гласных длительностью в 15, 20 мсек. и т. д. серьезно даже не обсуждался. По-видимому, трудности, которые необходимо было преодолеть в экспериментах по соотносению коротких отрезков гласных с фонемами, были настолько велики, что долгое время реальная возможность такого эксперимента ускользала от внимания исследователей.

Наши данные заставляют думать, что основная трудность оценки гласных такой малой длительности состоит не в возможном понижении точности слухового измерения спектральных характеристик этих звуков, а в том, что короткие отрезки гласных имеют групповой признак, отличающий их от естественных гласных, — они воспринимаются как щелчки. Наличие этого признака или тормозит на какое-то время, или полностью исключает существующие ассоциации между спектральными характеристиками сигнала и фонетическим качеством гласного. Сходный эффект наблюдался Горнбаллом (Turnbull, 1944) при попытках определить дифференциальные пороги по частоте для коротких тональных посылок методом сравнения их высоты с высотой тона значительной длительности. Уже при укорочении длительности посылки до 35—30 мсек. дифференциальные пороги оказались не определенными, так как сигналы относились к разным категориям. Вместе с тем дифференциальные пороги определялись без труда, если короткая посылка (щелчок) сравнивалась по высоте с другой посылкой такой же длительности (Oettinger, 1959; Лян Чжи-ань, Чистович, 1960; Cardozo, 1962).

Судя по нашим данным, переход гласных в категорию щелчков начинается при укорочении их длительности до 40 мсек. По литературным данным (Lübcke, 1921; Bürck et al., 1935; Türk, 1940), порог тональности на частоте 140—150 гц (основная частота гласных наших дикторов) составляет 23—25 мсек.

Выбор методики самообучения основывался на предположении, что восприятие коротких отрезков гласных как щелчков является помехой, затрудняющей использование существующих у человека ассоциаций между спектральными характеристиками сигнала и фонемами. Цель тренировки состояла в том, чтобы устранить это мешающее действие, приучив человека к тому, что в некотором частном случае искажений гласные могут звучать как щелчки.

Не сообщая аудитору сведений о правильности принятых им решений, мы надеялись тем самым исключить возможность образования новых ассоциаций, что соответствовало бы условиям «обучения с учителем».

Полученные результаты (необходимость очень длительной тренировки для достижения полной опознаваемости, большие индивидуальные различия между аудиторами, наличие провала в кривой опознаваемости в области 40 мсек.) говорят о том, что распространение опыта по опознаванию естественных гласных на категорию щелчков является очень трудной задачей, вероятно сравнимой по трудности с выработкой новых правил классификации, т. е. абсолютного опознавания. Это в какой-то мере перекликается с данными Либермана (Liberman, 1968) о том, что некоторые правила обработки сигнала, используемые при восприятии речевых сигналов, не применяются к близким по физическим характеристикам сигналам, воспринимаемым как неречевые звуки. Иными словами, выработанные правила классификации, очевидно, имеют локальное значение, действенное только для ограниченных подмножеств звуков.

ЗАКЛЮЧЕНИЕ

Теория информации считает, что естественная речь человека имеет огромную избыточность, которая обуславливает надежность ее приема и большую помехоустойчивость. Факт 100% опознаваемости коротких отрезков гласных, который получен у двух аудиторов в наших опытах, свидетельствует о том, что укорочение длительности этих отрезков до 15 мсек. не влияет на сохранность признаков, определяющих этот звук, как звук гласного речи.

В связи с этим же фактом и тем, что не все аудиторы обнаруживают способность к высокой опознаваемости в описанных выше условиях эксперимента, следует выдвинуть предположение, что наряду с существованием абсолютного музыкального слуха у человека существует и абсолютный тембральный слух. Последнее обстоятельство представляет немаловажную ценность для специалистов, занимающихся исследованием слуха.

Л и т е р а т у р а

- Л я н Ч ж и - а н ь, Л. А. Ч и с т о в и ч. 1960. Дифференциальные пороги по частоте в зависимости от длительности тональных посылок. Акуст. журн., 5 : 81—86.
- A h m e n d R., R. F a t e h c h a n d. 1959. Effect of sample duration on the articulation of sounds in normal and clipped speech. J. Acoust. Soc. Amer., 31, 7 : 1022—1029.
- B r a d y P. T., A. S. H o u s e, K. N. S t e v e n s. 1961. Perception of sounds characterized by a rapidly changing resonant frequency. J. Acoust. Soc. Amer., 33 : 1357—1362.
- B ü r c k W., P. K o t o w s k i, H. L i c h t e. 1935. Der Aufbau des Tonhöhenbewußtseins. Electr. Nachr. Techn., 12 : 326—337.

- Cardozo B. L. 1962. Frequency discrimination of the human ear. Intern. Congr. on Acoustics, Copenhagen : H-16.
- Heinz J. M., J. Lindqvist, B. F. Lindblom. 1967. Patterns of residual masking for sounds with speech-like characteristics. Confer. on Speech Communication a. processing, Boston : 246—251.
- Lieberman A. M. 1968. Models of speech perception. Speech Sympos., Kyoto, D-2 : 1—4.
- Lübcke E. 1921. Über Wahrnehmung kürzester Töne bei Unterwasserschallsendern. Ztschr. f. techn. Physik, 2 : 52—53.
- Oettinger R. 1959. Die Grenzen der Hörbarkeit von Frequenz- und Tonzahländerung. Acustica, 9 : 431.
- Turnbull W. W. 1944. Pitch discrimination as a function of tonal duration. J. Exper. Psychol., 34 : 302—316.
- Türk W. 1940. Über die physiologisch-akustischen Kennzeichen von Ausgleich vorgängen. Akust. Zschr., 5 : 129—142.

АКУСТИЧЕСКИЙ ПРИЗНАК ТВЕРДЫХ И МЯГКИХ СОГЛАСНЫХ РУССКОГО ЯЗЫКА В СВЯЗНОЙ РЕЧИ (СПЕКТРОГРАФИЧЕСКОЕ ИССЛЕДОВАНИЕ) ¹

*Шупляков В.,
Г. Фант, А. Серпа-Лейтао*

Институт физиологии им. И. П. Павлова АН СССР, Ленинград
Королевский технологический институт, Стокгольм

Различие между мягкими и твердыми согласными является особенностью системы согласных русского языка. Физиологическая основа такого различия заключается в наличии или отсутствии палатализации согласного. В терминологии дифференциальных признаков мягкие (палатализованные) и твердые (непалатализованные) согласные разделяются благодаря признаку дизный—простой,² акустической характеристикой которого является подъем F_2 и F_3 по мере того, как язык приближается к позиции, характерной для звука [i] (Fant, 1960).

Важность особого, [i]-образного перехода второй форманты для восприятия мягкости согласных неоднократно подчеркивалась в литературе (Дукельский, 1962; Зиндер и др., 1964). В работе

¹ Работа была выполнена в Лаборатории передачи речи Королевского технологического института в Стокгольме, руководимой проф. Г. Фантом.

² Согласно уточнению, предложенному в последнее время (Chomsky, Halle, 1968), палатализация рассматривается как наложение гласного [i] на комплекс признаков согласного и замена признака согласного дизный—простой признаками, характерными для гласных: высокий, низкий и задний.

Дукельского было показано, что при прослушивании слога, состоящего из мягкого согласного и гласного, и при постепенном отрезании начала согласного наступает момент, когда сам мягкий согласный перестает звучать, но мягкость продолжает восприниматься.

Несмотря на очевидную важность [i]-образного перехода для восприятия мягких согласных, вопрос о том, какие параметры перехода используются человеком при восприятии мягкости, является к настоящему времени неисследованным.

Можно сформулировать три равновероятные гипотезы относительно того, какую информацию о переходе человек использует при разделении согласных на твердые и мягкие в контексте.

1. Экстремальное положение F_2 всегда выше у мягкого согласного, чем у твердого, произнесенного в том же контексте и тем же диктором.

2. Все мягкие согласные имеют более высокую F_2 , чем все твердые, независимо от контекста:

а) это действительно для всех дикторов, независимо от размера речевого тракта;

б) это действительно только для одного диктора.

3. Вторая форманта мягкого согласного всегда выше, чем предшествующего и последующего гласного, т. е. переход положителен в последовательности ГС и отрицателен в последовательности СГ.

Целью настоящей работы была оценка этих гипотез на материале спектрограмм всех согласных русского языка в контексте с использованием в качестве дикторов как мужчин, так и женщин.

МЕТОДИКА

В экспериментах в качестве дикторов участвовали три женщины и двое мужчин русского происхождения. В задачу дикторов входило произнести таблицу из симметричных сочетаний типа ГСГ (гласный—согласный—гласный). Никаких ограничений на характер произношения не накладывалось. В таблице использовались все гласные и все согласные русского языка. Таким образом, имела возможность проанализировать все встречающиеся в русской речи переходы от согласного к гласному. Таблица состояла из 180 сочетаний. Записанные на магнитную ленту (магнитофон типа Ampex 261) слоги затем анализировались с помощью сонографа (Kay Electric Sonagraph typ 4691 A). Ширина полосы частот избирательного фильтра анализатора составляла 300 гц на всех частотах. При анализе использовалась частотная коррекция, заключающаяся в равномерном увеличении коэффициента усиления в сторону высоких частот на 6 дб на октаву, начиная с 1000 гц.

Измерения F_2 и F_3 производились на участке согласного в точках максимального или минимального значений, которые часто не совпадают с началом гласного (рис. 1). Если таковые отсутствовали (например, в случае взрывных), измерялось значение формант в самом начале перехода от согласного к гласному. Результаты измерений затем округлялись до ближайших 100 гц и откладывались на графике с учетом твердости или мягкости согласных.

На каждой спектрограмме отмечалось направление перехода.

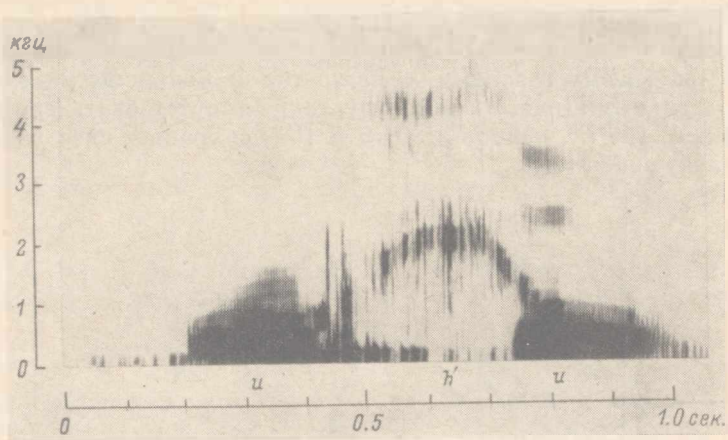


Рис. 1. Спектрограмма сочетания [uh'u].

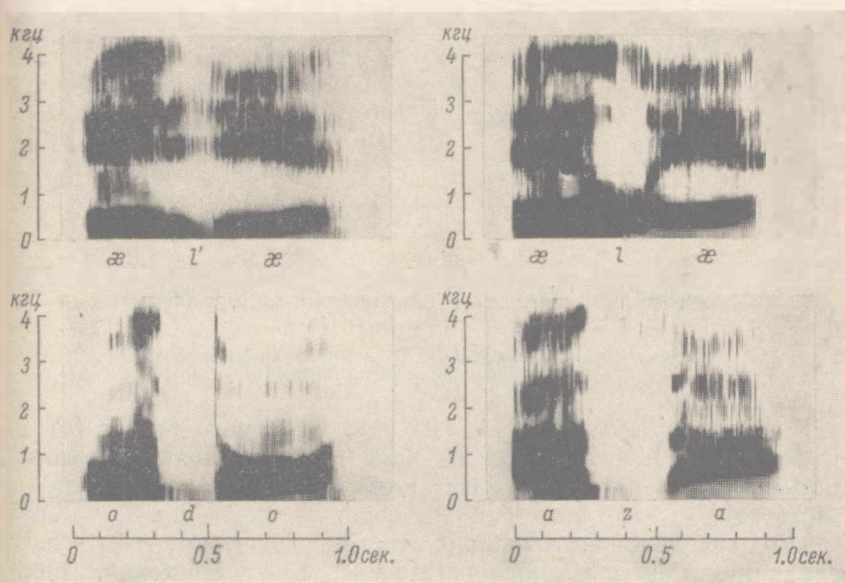


Рис. 2. Спектрограммы различных сочетаний.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Вскоре стало очевидно, что гипотеза 3 не может быть применена ко всему набору данных. Хотя переходы от мягкого согласного к последующему гласному и являются в большинстве случаев отрицательными (рис. 2), они могут также отсутствовать (см. [æl'æ] на рис. 1, где F_2 , равное 2100 гц в [l'], сохраняет свое значение в последующем [æ]).

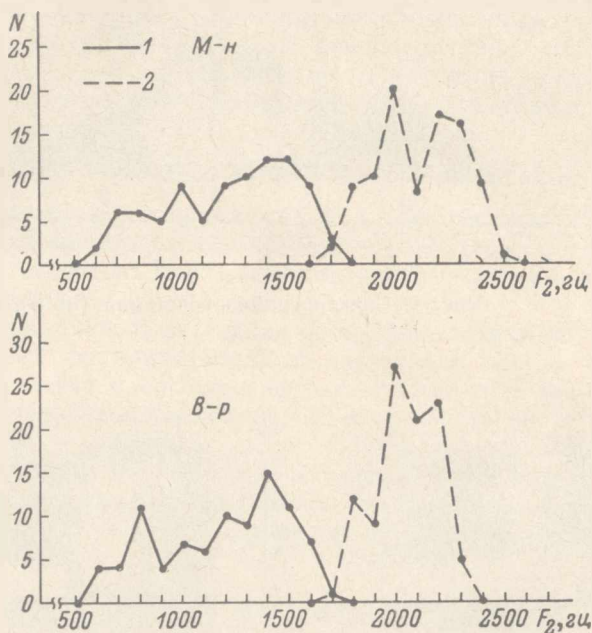


Рис. 3. Частота второй форманты у твердых (1) и мягких (2) согласных, произнесенных дикторами-мужчинами М-н и В-р.

На рис. 3 и 4: по оси абсцисс — частота второй форманты; по оси ординат — число случаев.

В последовательности [ælæ] на рис. 1 F_2 переходит от 830 гц на твердом [l] к 1900 гц на [æ] (переход положительный). В то же время в [odo] и [aza] переход отрицательный или отсутствует.

Таким образом, хотя направление перехода может быть важным параметром восприятия (Stevens, 1967; Чистович, 1968), оно не обеспечивает простого инвариантного правила для разделения согласных на твердые и мягкие.

Гипотезы 1 и 2 сохраняли свою значимость, поскольку F_2 мягкого согласного всегда оказывалась выше F_2 твердого в одинаковом контексте и у одного и того же диктора. Следующим

шагом требовалось оценить, имеется ли некоторый абсолютный критерий относительно F_2 , независимый от контекста, как это сформулировано в гипотезе 2.

Рис. 3 и 4 показывают распределение значений F_2 , с учетом мягкости или твердости согласных, для каждого диктора в отдельности. Видно, что кривые на рисунках очень мало перекрываются.

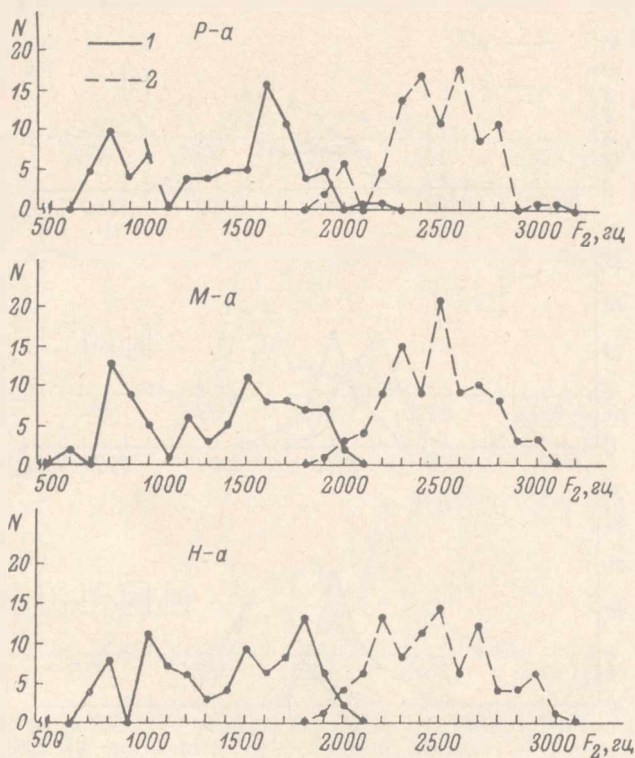


Рис. 4. Частота второй форманты у твердых (1) и мягких (2) согласных, произнесенных дикторами-женщинами Р-а, М-а и Н-а.

ваются. Точка взаимного пересечения находится на 1700 гц у обоих дикторов-мужчин и на 1900, 2000 и 2000 гц у дикторов-женщин. Число ошибок, которое получилось бы в случае применения этих пороговых значений ко всей совокупности произнесенных сочетаний, составляет у мужчин 1.9% из 360 реализаций, у женщин — 4% из 540. Анализ ошибок показывает, что из общего числа 29 ошибок 18 приходится на переходы в сочетаниях, не встречающихся в обычной русской речи или встречающихся очень редко, таких как [ætæ], [ihi], [uʃ'u], [Oʃ'O], [iʃ'i], [uɣ'u] и др. 17 ошибок были сделаны в сочетаниях с передними гласными.

Была сделана попытка оценить возможность уменьшения числа ошибок за счет информации о третьей форманте. Распределения значений F_3 для различных дикторов показаны на рис. 5—9. Как видно из рисунков, информация о третьей форманте согласного может быть эффективно использована лишь в случаях контекстов с передними гласными. При этом значения F_3 у мягких согласных

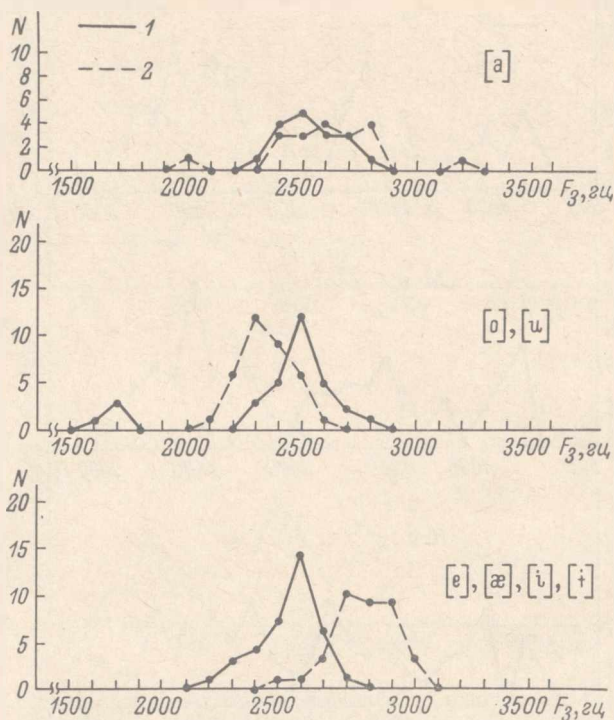


Рис. 5. Частота третьей форманты у твердых (1) и мягких (2) согласных, за которыми следуют различные гласные.
Диктор М—н.

На рис. 5—9: по оси абсцисс — частота третьей форманты;
по оси ординат — число случаев.

оказываются выше, чем у твердых. Аналогичная зависимость обнаруживается и в случае контекстов с гласными [o] и [u], однако у одного диктора (М—н) значения F_3 у мягких согласных оказались ниже, чем у твердых. Точки взаимного пересечения распределений F_3 для сочетаний с передними гласными находятся на 2600 гц у обоих дикторов-мужчин и на 3100, 3100 и 3000 гц у дикторов-женщин.

Таким образом, значения пороговых частот F_2 и F_3 , относительно которых могут приниматься решения о мягкости соглас-

ных, оказались различными у различных дикторов. Возникает вопрос, с какими общими характеристиками дикторов связаны значения этих порогов. Одним из параметров, характеризующих дикторов, может являться общая длина их артикуляционного аппарата, которая в первом приближении оказывается пропор-

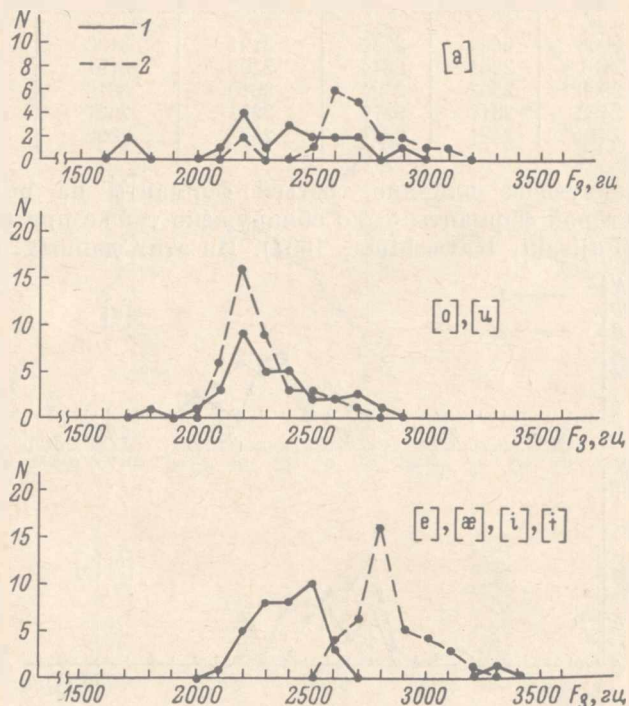


Рис. 6. Частота третьей форманты у твердых (1) и мягких (2) согласных, за которыми следуют различные гласные. Диктор В—р.

циональному среднему значению F_3 , измеренному при произнесении звука [i] (Fant, 1959). Были вычислены средние значения F_3 для различных гласных у каждого диктора; результаты вычислений представлены в таблице. Из таблицы видно, что с ошибкой около 3% можно допустить, что средние значения $\bar{F}_3$ в случае гласных [i] и [ɪ] совпадают с пороговым значением $F_{3\text{пор}}$. Пороговые значения F_2 оказались равными 0.53 от среднего значения F_3 в группе мужчин и 0.59 — в группе женщин. Таким образом:

$$F_{2\text{пор}} = 0.53 \bar{F}_3 \text{ (у мужчин),}$$

$$F_{2\text{пор}} = 0.59 \bar{F}_3 \text{ (у женщин),}$$

$$F_{3\text{пор}} = \bar{F}_3 \text{ (у мужчин и женщин).}$$

Средние значения третьей форманты (гц) для различных гласных у каждого диктора

Диктор	Гласный					Общее среднее
	[a]	[o]	[u]	[i], [ɨ]	[e], [æ]	
М—а	3008	2844	2780	3191	3132	2991
Н—а	2181	2961	2844	3208	3191	3097
Р—а	2800	2532	2524	3010	3016	2776
В—р	2492	2317	2277	2657	2621	2473
М—н	2578	2421	2321	2602	2622	2509

Нормализующее влияние третьей форманты на положение границы второй форманты было обнаружено также при изучении гласных (Fujisaki, Kawashima, 1967). Из этих данных, а также

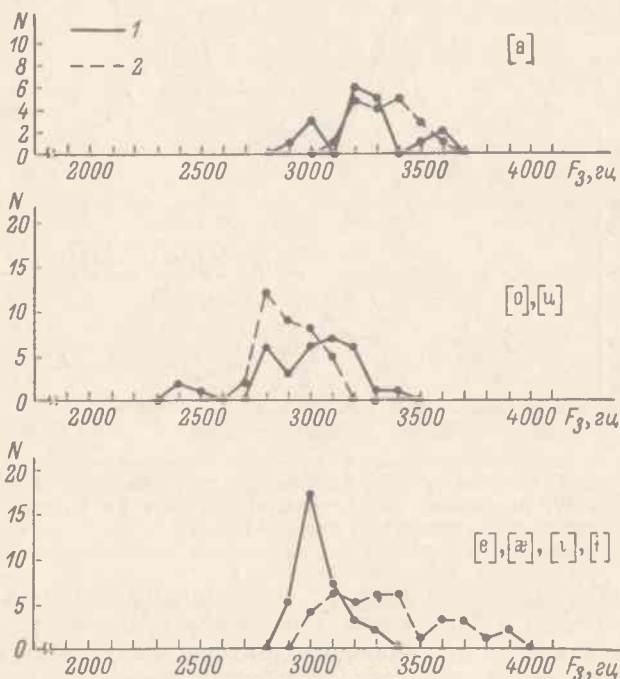


Рис. 7. Частота третьей форманты у твердых (1) и мягких (2) согласных, за которыми следуют различные гласные. Диктор Н—а.

из работы Фанта (Fant, 1959) следует, что F_3 оказывается особенно важной для идентификации передних гласных.

Обращает на себя внимание очень малый разброс пороговых значений формант среди групп мужчин и женщин. Возможно, сведения

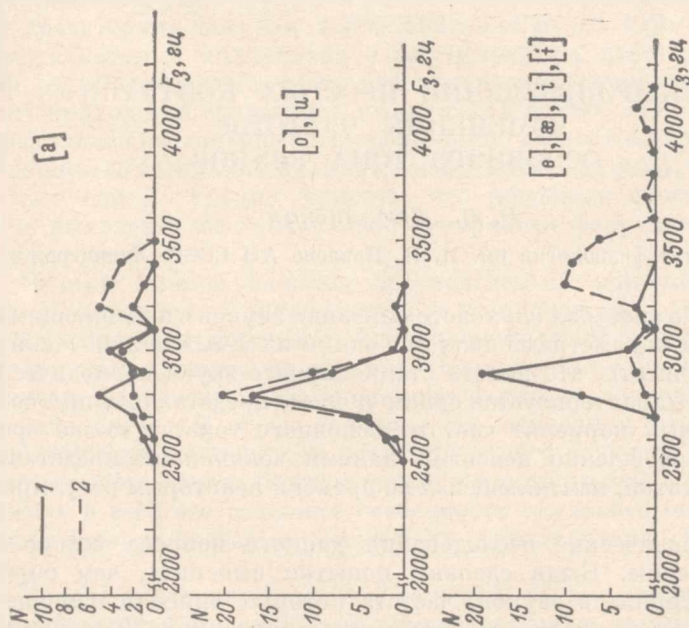


Рис. 8. Частота третьей форманты у твердых (1) и мягких (2) согласных, за которыми следуют различные гласные. Диктор М—а.

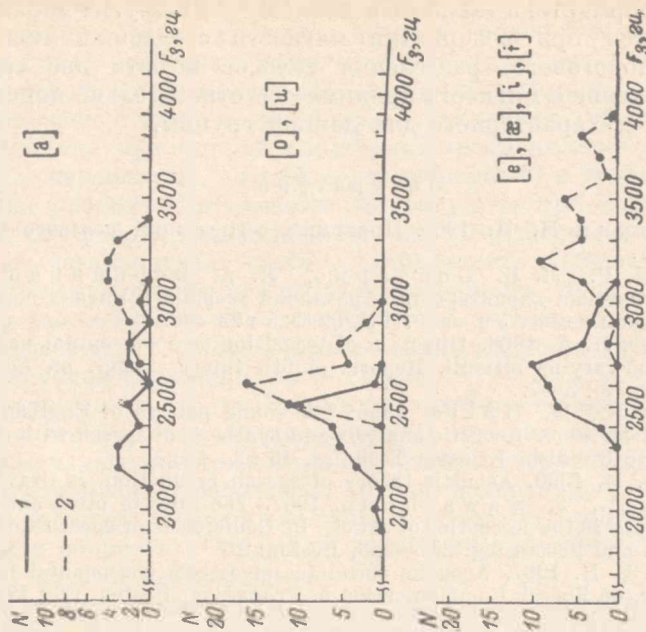


Рис. 9. Частота третьей форманты у твердых (1) и мягких (2) согласных, за которыми следуют различные гласные. Диктор Р—а.

о третьей форманте, а также об основном тоне и других параметрах используются при восприятии мягкости согласных лишь для грубого, порогового, разделения голосов на эти две группы, а затем решение о мягкости принимается относительно порогового значения F_2 , характерного для данной группы.

Л и т е р а т у р а

- Дукельский Н. И. 1962. Принципы сегментации речевого потока. М.—Л.
- Зиндер Л. Р., Л. В. Бондарко, Л. А. Вербницкая. 1964. Акустическая характеристика различия твердых и мягких согласных в русском языке. Уч. зап. ЛГУ, 69, 325 : 28—36.
- Чистович Л. А. 1968. Direction of transition as a perceptual parameter of time-varying stimuli. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2 : 3—7.
- Chomsky N., M. Halle. 1968. The sound pattern of English. N. Y.
- Fant G. M. 1959. Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics, 15 : 3—108.
- Fant G. M. 1960. Acoustic theory of speech production. 's-Gravenhage.
- Fujisaki H., T. Kawashima. 1967. The roles of pitch and higher formants in the perception of vowels. In: Conference on speech Communication and Processing: 251—256, Boston.
- Stevens K. N. 1967. Acoustic correlates of certain consonantal features. Confer. on Speech Communication a. Processing, Boston, C6 : 177—184.

ВОСПРОИЗВЕДЕНИЕ ПРОСТЫХ КОНТУРОВ ИЗМЕНЕНИЯ ЧАСТОТЫ ОСНОВНОГО ТОНА ЗВУКОВ

В. В. Люблинская

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

Вопрос о способах слухового описания звуков с изменяющимися во времени параметрами является одним из самых малоизученных. Принято считать, что высота стационарного звука периодической структуры характеризуется одним числом, представляющим собой субъективный коррелят частоты основного тона, в то же время остается совершенно неясным, какими величинами представлен звук с частотой, изменяющейся во времени некоторым регулярным способом.

Психофизические исследования данного вопроса весьма немногочисленны. Были сделаны попытки выяснить, чем определяется восприятие звуков, частота которых линейно изменяется на протяжении короткого промежутка времени — 20 и 50 мсек. (Brady et al., 1961; Heinz et al., 1967). В первой из цитируемых

работ использовались звуки, представляющие собой отклик резонансного фильтра, возбуждаемого периодической последовательностью коротких импульсов при линейном изменении его центральной части, во второй — колебания чистого тона линейно изменяющейся частоты. В обоих случаях была применена процедура опыта, при которой испытуемым предъявлялись два стимула (один — переменный, другой — стационарный) и требовалось оба стимула уравнивать по «высоте», варьируя частоту стационарного звука. Обе работы показывают, что при выполнении такой задачи человек устанавливает частоту стационарного звука, в среднем очень близкую к конечной частоте переменного. Из этого факта делается заключение о том, что конечная частота монотонно изменяющегося звука (ее субъективное измерение) есть та особая величина, которая характеризует его высоту.

Кроме того, в работе Брэди и других была обнаружена тенденция к экстраполяции оценки конечного значения частоты в направлении ее изменения, что послужило основанием для предположения об осуществлении человеком других измерений, описывающих как-то характер движения частоты во времени (например, траекторию или скорость). Однако во второй из приведенных работ этот эффект не был подтвержден.

Естественно полагать, что звуки с изменяющимися во времени параметрами в отличие от стационарных звуков представлены не одной величиной — характеристическим измерением одной точки на траектории, так как такое описание будет явно неполным. Метод, который применялся в рассмотренных работах, обладает тем серьезным недостатком, что в условиях эксперимента человеку приходится сравнивать между собой звуки, один из которых представлен несколькими измерениями по исследуемому параметру (в данном случае таким параметром является «высота»), а другой — только одним. Трудно ожидать, что подобным способом могут быть выявлены все субъективно измеряемые величины нестационарных стимулов.

В этом смысле кажется перспективным использовать процедуру имитации — артикуляторное воспроизведение звуков с заданным законом изменения его параметров. Процесс имитации организован следующим образом. Воспринимая звуковой сигнал, человек измеряет и запоминает определенные величины, описывающие слуховой образ сигнала. На основании этих величин формируются команды для артикуляторного осуществления соответствующего звука. Характеристики ответного звука должны отражать в себе все основные особенности слухового описания исходного звука, которые могут быть артикуляторно воспроизведены.

Исходя из приведенных соображений, была выполнена работа разведывательного характера, имевшая целью выяснить, какие существенные измерения звука с изменяющейся во времени

частотой основного тона будут сохраняться при его воспроизведении.

Конкретное изучение данного вопроса определяется следующим обстоятельством. Артикуляторная система человека обладает инерционностью, обуславливающей значительное сглаживание быстрых изменений исходного звука при его воспроизведении. Например, при имитации сигнала, частота которого в определенный момент изменяется скачками на некоторую ощутимую величину, человек будет создавать сигнал с плавным изменением частоты, обуславливаемым переходным процессом в системе артикуляции.

На основании существования инерционности артикуляторной системы некоторые исследователи предполагают, что плавные изменения частоты основного тона, образующие интонацию речевого высказывания, на уровне исходного субъективного описания могут быть заданы как скачкообразные (Lieberman, 1967; Ohman, 1967).

При исследовании способа описания изменяющихся сигналов с помощью метода имитации следует четко представлять, какие свойства воспроизведенных человеком звуков отражают особенности его слухового образа, а какие — те искажения, которые вносятся при его реализации. Для того чтобы иметь представление об инерционных свойствах голосообразующей системы, в настоящих экспериментах в качестве стимулов использовались наряду со звуками, частота основного тона которых изменялась монотонно и достаточно медленно, такие звуки, частота которых изменялась скачком на заданную величину от f_1 до f_2 . При этом предполагалось, что вследствие малой инерционности слуховой системы — существенно меньшей, чем инерционность артикуляторно-двигательного аппарата, — подобные сигналы в их субъективном описании сохраняют свою скачкообразную структуру, т. е. заданы в виде двух величин — измерений частоты f_1 и f_2 .

Исходным сигналом в экспериментах служил звук [а], синтезированный с помощью формантного синтезатора, частота голосового источника которого могла изменяться по заданному закону. Частота основного тона на протяжении каждого звука изменялась по одной из трех траекторий, форма которых схематически приведена на рис. 1.

I тип стимулов представлял собой переменный звук, траектория изменения частоты основного тона которого имела форму линейного склона на частотно-временной плоскости. Длительность склона 300 мсек. II тип стимулов отличался от I типа наличием стационарного участка длительностью 200 мсек., предшествующего склону с теми же характеристиками, что и для стимулов I типа. III — тип стимулов — ступенька — состоял из двух стационарных отрезков одинаковой длительности (150 мсек.), различающихся по частоте. Все стимулы имели восходящий характер изменения частоты основного тона. Начальная частота во всех случаях была 200 гц, а конечная варьировала от 240 до 400 гц ступенями примерно в 40 гц. За исключением частоты основного тона, все остальные параметры синтезатора были инвариантны.

Стимулы в тестовой последовательности имели следующий порядок: сначала предъявлялись стимулы I типа, варьующие от максимального изменения

частоты до минимального, затем стимулы II и далее III типа с тем же характером варьирования. Каждый стимул повторялся трижды.

Испытуемый, находившийся в звукозаглушенной камере, слушал звуки и последовательно воспроизводил каждый, стараясь имитировать его интонацию. Горловым микрофоном регистрировались ответные голосовые колебания испытуемого, частота основного тона которых измерялась с помощью периодометра. В опытах участвовали трое испытуемых с нормальным слухом, не имеющие специального музыкального образования, но способные довольно хорошо воспроизводить голосом абсолютные значения частоты одиночных тонов в том диапазоне частот, в котором варьировали настоящие стимулы (Lublinskaja, 1968).

На рис. 2 представлены типичные периодограммы звуков, созданных при имитации стимулов всех трех типов. Как можно

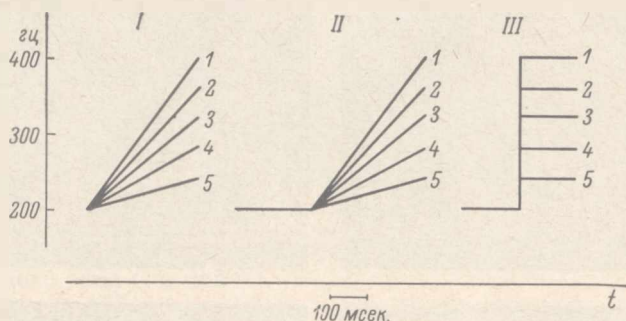


Рис. 1. Форма траекторий изменения частоты основного тона стимулов.

По оси абсцисс — время, мсек.; по оси ординат — частота, гц.
Римские цифры — типы стимулов, арабские — варианты стимулов одного типа.

видеть, при имитации звуков с плавным изменением частоты основного тона (стимулы I и II типов) воспроизведенная траектория изменения частоты основного тона в своей изменяющейся части в значительной степени подобна исходной (рис. 2, а—г). Однако ответные сигналы имеют существенную особенность. В начале и конце склона значительного числа ответных периодограмм, соответствующих стимулам I типа, присутствуют стационарные участки (рис. 2, б). Их длительность невелика (порядка 20—100 мсек.) и заметно меньше длительности стационарных участков тех ответных звуков, исходные стимулы которых содержат стационарные отрезки частоты основного тона (стимулы II и III типов). В конце звука стационарные участки наблюдались в большинстве случаев, а в начале — менее регулярно (рис. 2, в). Такое же завершение характерно для периодограмм имитации стимулов II типа (рис. 2, г).

Наблюдаемая структура имитирующих сигналов позволяет думать о том, что субъективное описание интонационной картины звука представляет собой набор существенных измерений —

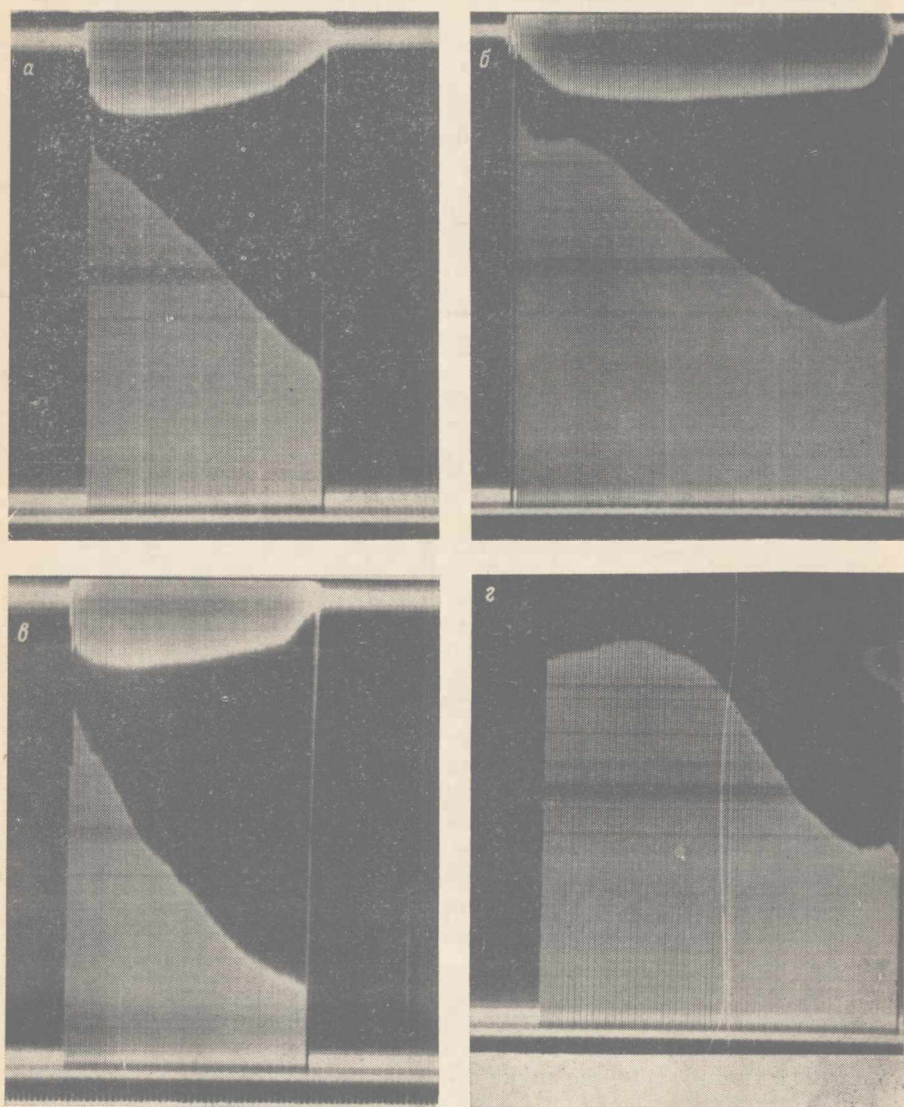


Рис. 2. Перидограммы ответных звуков.

а, б, в — имитирующих стимулы I типа; *г* — II типа; *д* — III типа. Внизу — отклонение луча, пропорциональное периоду основного тона звука; наверху — колебания регистрируемые горловым микрофоном,

характеристических величин, а не сводится к полному поточечному ее изображению, получаемому с помощью регулярных во времени отсчетов, как это предполагается в известной модели «анализа через синтез» (Heinz, 1962).

Таковыми характеристическими величинами среди прочих, очевидно, являются измерения частоты в начале и конце склона, что и обуславливает некоторую фиксацию состояний голосовых связок в соответствующих положениях.

Интересно было выяснить, насколько точно воспроизводятся краевые значения изменяющейся части траектории частоты основного тона. Для этого в ответных звуках измерялись начальная (или установившаяся) частота основного тона f_n и конечная частота f_k . На периодограммах рис. 3 отмечены

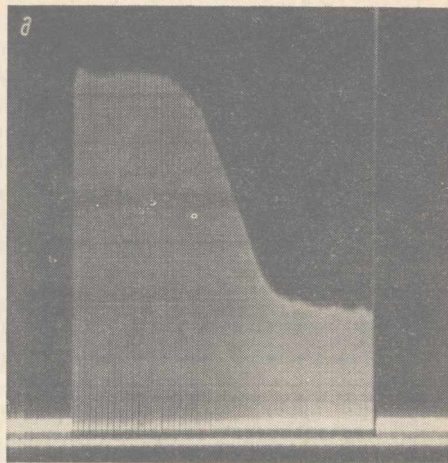


Рис. 2. (продолжение).

соответствующие отсчеты $T_n = \frac{1}{f_n}$ и $T_k = \frac{1}{f_k}$. Полученные в результате этих измерений данные представлены на рис. 4.

Приведенные данные убедительно свидетельствуют о том, что человек достаточно точно воспроизводит краевые значения

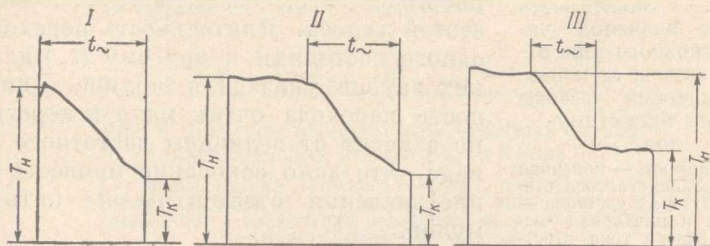


Рис. 3. Контуры периодограмм ответных звуков, имитирующих стимулы I, II и III типов.

По оси абсцисс — время; по оси ординат — отклонение луча, пропорциональное периоду основного тона. Остальные объяснения в тексте.

частоты основного тона (как начальные, так и конечные) при монотонном ее изменении. Можно также видеть, что воспроизведение краевых значений склона довольно мало отличается от воспроизведения частот стационарных участков стимулов при скач-

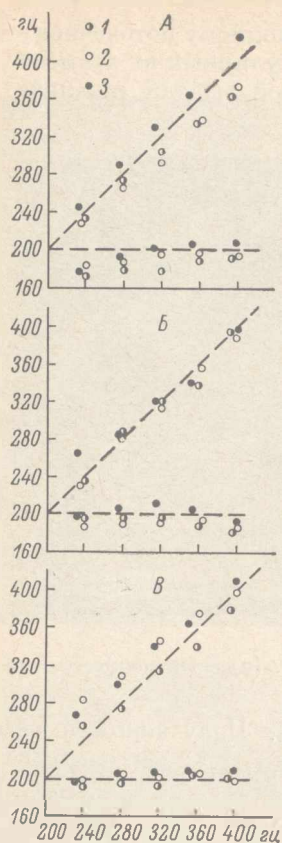


Рис. 4. Зависимость краевых значений частоты основного тона ответных звуков от конечного значения частоты основного тона стимулов.

По оси абсцисс — конечная частота основного тона стимулов, гц; по оси ординат — конечная и начальная частота основного тона ответных звуков, гц. А, Б, В — данные испытуемых Р. Ш., В. Л. и Н. С. соответственно. Измерения в ответах на стимулы: 1 — I типа, 2 — II типа, 3 — III типа. Прерывистые линии — совпадение конечных значений частоты основного тона ответных звуков с соответствующими измерениями стимулов (наклонная прямая относится к конечным значениям, горизонтальная — к начальным).

кообразном изменении частоты. В первом случае, как правило, конечная частота несколько меньше, а начальная — несколько больше частоты стационарных звуков.

Настоящие результаты подтверждают предположение о том, что краевые значения служат существенными измерениями, необходимыми для описания изменяющегося сигнала. То, что человек каким-то образом измеряет конечную частоту склона, как уже было сказано выше, отмечалось рядом авторов. Новым результатом является установление факта измерения начального значения.

Более интересно, конечно, выяснить, как оценивается информация о характере самого изменения частоты. Достаточно ли для описания звуков I и II типов задания двух величин, соответствующих началу и концу склона, или имеются другие измерения? Ответ на этот вопрос можно получить, сопоставив ответы на стимулы I, II и III типов. Для описания последних достаточно двух измерений частоты основного тона, тем не менее, как можно видеть (рис. 2, д), ответные звуки на эти стимулы также содержат участок плавного изменения частоты, что явно обусловлено инерционностью системы управления частотой голоса. Длительность перехода от одного состояния к другому ($t_{\text{переход}}$) для таких звуков указана в таблице. Длительность перехода очень мало и нерегулярно зависит от величины частотного перепада, что дало основание привести средние значения ответов на все пять стимулов.

В той же таблице указаны длительности переменного участка периодограмм ответов на стимулы I и II типов. Можно видеть, что длительность переменного участка ответных звуков, имитирующих плавные изменения частоты основного тона, существенно превышает переходный период при имитации скачка. Этот факт свидетельствует о том, что изменение ча-

Длительность переменного участка (средние значения $\bar{t}$, мсек.)
и стандартное отклонение (σ , мсек.) периодограмм ответных звуков

Тип стимула	Испытуемые					
	Р. Ш.		В. Л.		Н. С.	
	$\bar{t}$	σ	$\bar{t}$	σ	$\bar{t}$	σ
I	295.5	56.3	365.0	93.5	303.8	84.8
II	381.3	118.2	340.9	66.6	360.2	65.4
III	217.4	56.4	244.4	40.3	229.6	57.0

стоты в первом случае определяется не переходным режимом и, значит, описание этих звуков не ограничивается двумя крайними измерениями, а имеется определенное указание о характере изменения.

Последнее подтверждается также при качественном сравнении между собой формы периодограмм ответов на стимулы I и II типов

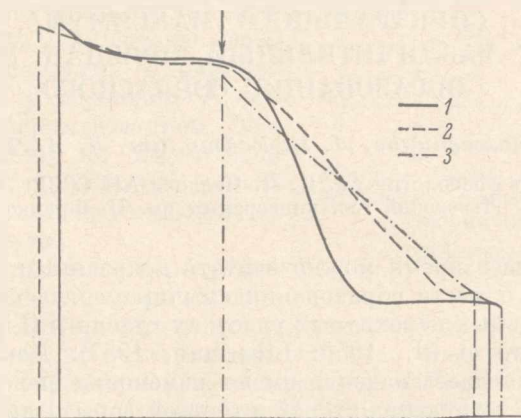


Рис. 5. Контуры периодограмм ответных звуков, имитирующих стимулы трех типов.

1 — периодограмма ответного сигнала, имитирующего стимул III типа; 2 — II типа; 3 — I типа. Стрелкой указана точка совмещения периодограмм, соответствующая началу изменения частоты.

с периодограммами ответов на стимулы III типа (рис. 5). Различие в форме тректорий очевидно.

К сожалению, на основании чисто качественного анализа невозможно определенно сказать, по каким признакам описывается характер изменения сигналов. Конкретное выяснение этого вопроса может быть осуществлено при строгом количественном анализе формы интонограмм воспроизведенных звуков, что явится предметом дальнейших исследований.

Л и т е р а т у р а

- Brady P. T., A. S. House, K. N. Stevens. 1961. Perception of sounds characterized by a rapidly changing resonant frequency. *J. Acoust. Soc. Amer.*, 33 : 1357—1371.
- Heinz J. M. 1962. An analysis of speech spectra in terms of a model of articulation. *Speech Communication Seminar, Stockholm, sess. B2.*
- Heinz J. M., B. E. F. Lindbloom, J. Ch. K.-G. Lindqvist. 1967. Pattern of residual masking for sounds with speech-like characteristics. *Confer. on Speech Communication a. Processing, Boston* : 246—251.
- Lieberman P. 1967. *Intonation, perception and language.* Cambridge, Massachusetts.
- Lublinskaja V. V. 1968. Mimicking of pitch. *Zschr. f. Phonetik*, 21 : 129—134.
- Öhman S. 1967. Word and sentence intonation: a quantitative model. *STL—QPSR*, 2—3 : 20—54.

ЗНАК ИЗМЕНЕНИЯ ЧАСТОТЫ СПЕКТРАЛЬНОГО МАКСИМУМА КАК РАЗЛИЧИТЕЛЬНЫЙ ПРИЗНАК МЕСТА ОБРАЗОВАНИЯ СОГЛАСНОГО

В. А. Кожевников, Л. Е. Рейтблат, Л. А. Чистович

Институт физиологии им. И. П. Павлова АН СССР, Ленинград;
Львовский гос. университет им. И. Франко

В настоящее время можно считать доказанным, что основная информация о месте образования смычных согласных заключена в так называемых переходных участках гласного (Liberman et al., 1954; Delattre et al., 1955; Liberman, 1957). Доказано также, что преобладающее значение имеют изменения спектра в области частот, соответствующих второй и третьей формантам (Harris et al., 1958; Hoffman, 1958). Значительно менее ясен вопрос о том, в чем конкретно содержится эта информация и с помощью каких механизмов она извлекается слуховой системой.

Исследуя зависимость идентификации синтетических стимулов от характеристик перехода второй форманты, Делатр и др. (Delattre et al., 1955) выдвинули гипотезу «локусов». Хотя эта гипотеза обычно рассматривается как относящаяся к восприятию, она имеет достаточно отчетливый смысл лишь в плане выбора параметров переходов формант при синтезе согласных.

Делатр и др. (Delattre et al., 1955) предложили следующую схему синтеза, поясняемую рис. 1. На основании согласной фонемы выбирается одно из трех возможных начальных значений второй форманты (F_2). Это значение и названо авторами «локусом». Обозначим его F_1 (F_{1_1} , F_{1_2} , F_{1_3} . .). На основании гласного выби-

рается значение F_2 на стационарном участке ($F_{s1}, F_{s2}, F_{s3} \dots$). Если закон изменения F_2 во времени задан и задано также полное время перехода (T), то величины F_i и F_s однозначно определяют как направление, так и скорость перехода. Однако эксперименты показали, что необходимо еще задать и реальную длительность перехода (компонента сигнала, соответствующая переходному участку F_2 , должна быть выраженной лишь на отрезке времени D , меньшем T). Кроме того, оказалось, что само значение F_i следует выбирать не только на основании согласного, но и с учетом гласного (Delattre et al., 1955; Liberman et al., 1967).

Никаких данных в пользу того, что при распознавании согласных нервной системой человека производится «вычисление» местоположения F_i , получено не было. Более того, легко заметить, что для вычисления значения F_i , принятого при синтезе, слушателю необходимо располагать априорной информацией о значении T , выбранном экспериментатором. Так как он заведомо не имеет этой информации, то гипотеза о восприятии локуса неизбежно должна быть отвергнута.

Если обратиться непосредственно к исходным экспериментальным данным Либермана и др. (Liberman et al., 1954), то можно заметить очень простую закономерность. Она состоит в том, что сигналы, характеризующиеся подъемом F_2 во времени, определяются как слоги, начинающиеся с губного согласного [p], [b]. При понижении же F_2 выбирается язычный согласный [t], [d] или [k], [g]. Это наводит на мысль, что слуховая система определяет направление изменения частотного положения спектрального максимума и что именно этот признак используется для классификации гласных на губные и язычные.

МЕТОДИКА

В настоящей работе мы исходили из предположения, что если человек способен определять направление сдвига спектрального максимума, то он должен делать это и тогда, когда частота максимума изменяется ступенькой, т. е. когда речеподобный звук представляет собой последовательность двух простых сегментов. Моделирование слога взрывной согласный — гласный двумя сегментами (посылка шума и стационарный гласный) производилось в работе Купера и др. (Cooper et al., 1952). Рассмотрение полученных ими данных действительно позво-

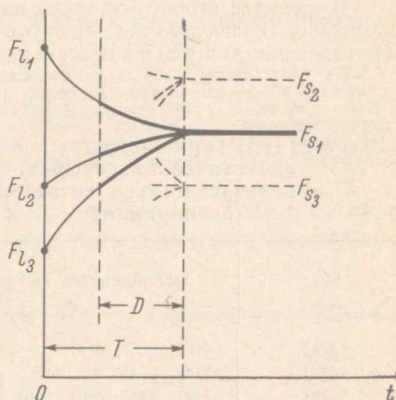


Рис. 1. Схема синтеза второй форманты, основывающаяся на теории локусов.

Объяснение в тексте.

ляет предполагать, что испытуемые ориентировались не столько на абсолютные значения частоты шума, сколько на разность частот шума и спектрального максимума гласного.

Для проверки предположения о роли знака изменения (направлении сдвига) спектрального максимума как признака, разделяющего губные и язычные согласные, целесообразно воспользоваться методом активного поиска фонетических границ (Чистович, 1968). Если устанавливаемая испытуемым граничная частота максимума на первом сегменте звуковой последовательности будет совпадать с частотой максимума на втором сегменте, то высказанная гипотеза подтвердится.

В качестве второго сегмента нами был взят упрощенный гласноподобный комплекс. В качестве же первого сегмента использовалась короткая посылка фильтрованного шума с шириной полосы 20%.

Таблица 1

Частота (гц) гармонических компонент
синтетических гласных,
использовавшихся во втором
эксперименте

$F_{гг}$	Ослабленные гармоники
300	100, 200, 400
400	200, 600
480	120, 240, 360, 600
600	150, 300, 450, 750
750	150, 300, 450, 600, 900
800	200, 400, 600, 1000
900	300, 600, 1200
1000	250, 500, 750, 1250
1200	300, 600, 900, 1500
1400	280, 560, 840, 1120, 1680
1500	300, 600, 900, 1200, 1800
1700	340, 680, 1020, 1360, 2040
1800	300, 600, 900, 1200, 1500

тах было замечено, что путем уменьшения относительной интенсивности и длительности этой посылки можно добиться того, что испытуемые воспринимают только [p] или [t]. Для нас этот эффект оказался удобным, так как позволил ограничиться поиском границы только между [p] и [t].

Блок-схема установки, служившей для создания звуковых сигналов, показана на рис. 2. Гласноподобный сегмент получался с помощью генератора сложных напряжений типа ГСН-1, способного создавать сигнал, состоящий из основного синусоидального колебания любой звуковой частоты и ряда высших гармонических компонент (до 6-й) желаемой амплитуды. В первой серии опытов применялись два рода сигналов, воспринимавшихся как [u] и [a]. [И]-образный гласный

состоял всего из двух гармоник: 300 и 600 гц. Уровень второй гармоники был на 13 дБ ниже первой. [a]-образный гласный состоял из первой гармоники 300 гц и гармоник 900 и 1200 гц примерно одинаковой амплитуды.

Характеристики гласных, использовавшихся во втором эксперименте, приведены в табл. 1. Уровень одной гармоники в этих стимулах был всегда на 20 дБ выше остальных. Ее частота принималась за частоту спектрального максимума второго сегмента ($F_{гг}$). Из табл. 1 видно, что как основная частота гласного, так и общее число гармоник в его спектре варьировали очень сильно.

В первом эксперименте в качестве первого сегмента звукового сигнала применялась посылка полосового шума с двумя фиксированными длительностями 10 и 15 мсек. Во втором эксперименте испытуемому разрешалось самому подбирать не только уровень интенсивности, но и длительность шума (в пределах 10—15 мсек.) так, чтобы максимально уменьшить вероятность восприятия [k].

В опытах испытуемый сам поворачивал ручку настройки частоты фильтра Φ (рис. 2), в качестве которого использовался селективный милливольтметр типа В2-6. Лимб со шкалой частот был невидим испытуемому. Частота, с которой на-

чинался поиск границы между [p] и [t], устанавливалась на крайних самых высоких или самых низких значениях шкалы фильтра. Значением частоты шума считалась частота максимума характеристики фильтра, имеющей вид резонансной кривой с шириной полосы 20%.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В первой серии опытов приняли участие 5 испытуемых. Объединенные данные по всем испытуемым и обеим длительностям шумовой послышки приведены на рис. 3. Рисунки показывают гistogramмы значений границ между [p] и [t], установленных в «слоге» с гласным [u] и [a]. Можно видеть, что распределения между со-

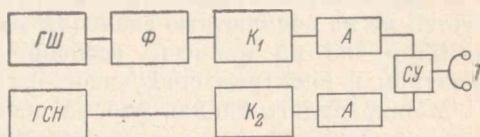


Рис. 2. Блок-схема экспериментальной установки.

ГШ — генератор шумовых колебаний; ГСН — генератор сложных напряжений; Ф — фильтр, перестраиваемый испытуемым; K_1 и K_2 — электронные ключи со схемами, задающими время открытия; А — аттенюаторы; СУ — суммирующее устройство; Т — динамические телефоны.

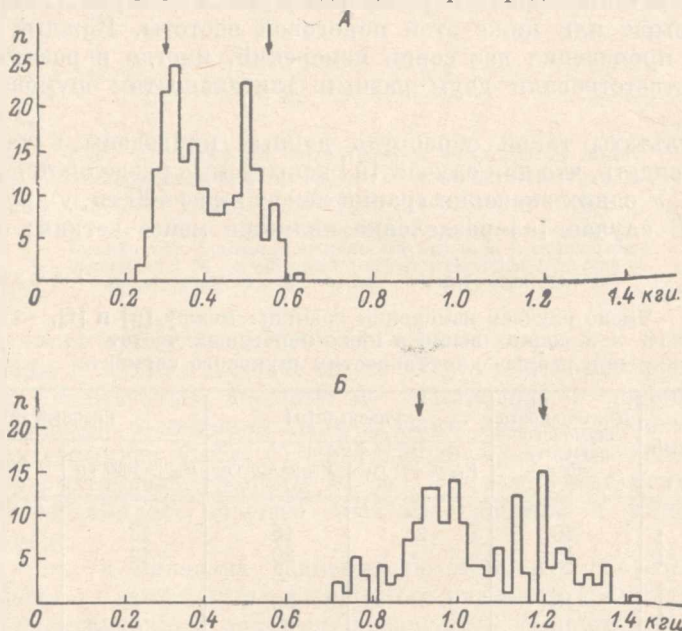


Рис. 3. Гistogramмы установленных значений границ между [p] и [t].

А — в слоге с гласным [u]; Б — в слоге с гласным [a]. По оси абсцисс — установка частоты фильтра, выделяющего полосу шума, кГц; по оси ординат — число случаев. Стрелки показывают значения частоты двух спектральных составляющих гласных звуков.

бой не пересекаются. Основная часть значений установленных границ в каждом из распределений заключена на интервале шкалы,

ограниченном частотами составляющих в спектре гласного (указано стрелками). Подобный характер данных доказывает, что при определении границы между [p] и [t] испытуемый ориентируется не на абсолютные значения максимума в спектре шумовой посылки, но на какие-то особенности отношения между этим спектром и спектром следующего за шумом гласного.

Можно видеть также, что на обеих гистограммах достаточно четко выделяются по два максимума, один из которых близок по частоте к более низкой, а другой — к более высокой из частотных составляющих гласного. Возникает предположение, что испытуемые выбирают в качестве спектрального максимума гласного или одну, или другую гармонику и именно с ней сравнивают частоту (спектральный максимум) шума. Чтобы проверить это предположение, были выбраны две частоты, примерно соответствующие минимуму, лежащему между пиками на гистограммах. Для слога с [u] была принята частота 420 гц, для слога с [a] — 1060 гц. Определялось число случаев, когда значения границ, установленных каждым испытуемым в серии измерений, лежат выше или ниже этой пороговой частоты. Каждый испытуемый производил две серии измерений, иногда в разные дни, они соответствовали двум разным длительностям шумовой посылки.

Результаты такой обработки данных приведены в табл. 2. Можно видеть, что для случая [u] испытуемые разделяются на две группы. У одних значения границ лежат ниже 420 гц, у других — выше. В случае [a] разделение является менее четким; однако

Таблица 2

Число случаев измерений границы между [p] и [t],
лежащих выше и ниже пороговых частот
при разных длительностях шумового сегмента

Испытуемый	Длительность шумового сегмента, мсек.	Гласный [u]		Гласный [a]	
		$F_{ш} \leq 420$ гц	$F_{ш} > 420$ гц	$F_{ш} \geq 1060$ гц	$F_{ш} > 1060$ гц
Л. Р. {	10	2	18	13	12
	15	—	20	12	3
А. Н. {	10	—	20	3	7
	15	—	20	6	4
М. Ж. {	10	19	1	—	20
	15	20	—	19	1
Ю. К. {	10	15	5	14	—
	15	20	—	15	—
В. К. {	10	20	—	—	20
	20	20	—	20	—

любопытно, что один и тот же испытуемый в одной серии измерений устанавливает все границы ниже 1060 гц, а в другой серии — выше 1060 гц (испытуемые В. К. и М. Ж.).

Таким образом, предположение о конкуренции двух максимумов на спектре гласного кажется довольно правдоподобным. Исходя из данных Пломпа (Plomp, 1964) и Пломпа и Мимпена (Plomp, Mimpen, 1968), необходимо допустить, что на слуховых эквивалентах спектров применявшихся нами гласных должны быть отдельные максимумы, соответствующие гармоникам.

Для того чтобы избежать неопределенности в отношении спектрального максимума гласного, в следующей серии опытов

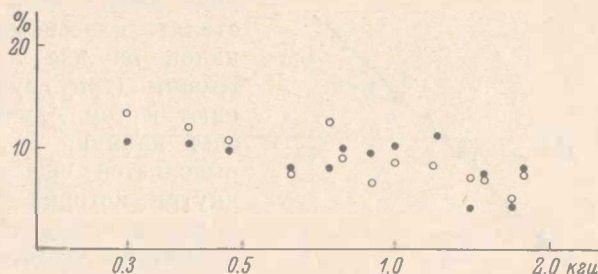


Рис. 4. Величина стандартного отклонения значений границы между [р] и [т] в зависимости от частоты спектрального максимума гласного.

По оси абсцисс — частота спектрального максимума используемого гласного, кГц; по оси ординат — стандартное отклонение, % от среднего значения границы. Кружки — испытуемый Л. Ч., точки — испытуемый В. К.

применялись стимулы, в которых одна из гармоник превышала по уровню остальные не менее чем на 20 дБ. Опыты были поставлены на двух испытуемых, которые не участвовали в первой серии экспериментов. Измерения были сделаны при 13 различающихся гласных стимулах (табл. 1). Для каждого гласного испытуемый повторял установку границы 20 раз. На основании этих 20 величин вычислялось среднее значение границы и стандартное отклонение.

На рис. 4 показана зависимость стандартного отклонения (выраженного в процентах от среднего значения границы) от частоты спектрального максимума гласного. Можно видеть, что стандартное отклонение в большинстве случаев не превышает 10%. Это сравнительно небольшая величина, если учесть, что по данным Фланагана (Flanagan, 1955), дифференциальный порог по частоте форманты, определенный для стационарного сигнала, составляет 3—5%. Из рис. 4 следует, что имеется некоторая тенденция уменьшения стандартного отклонения с увеличением абсолютного значения частоты.

На рис. 5 показана зависимость среднего значения границы от частоты спектрального максимума гласного. Можно видеть, что граница совпадает с частотой максимума гласного или находится достаточно близко от него. Это согласуется с предположением, что человек использует в качестве различительного признака знак разности между значениями частот шума и гласного, т. е. направление изменения во времени положения спектрального максимума сигнала.

Система, производящая определение знака (направления) изменения частоты, естественно, должна включать в себя пороговый элемент. Она по самой природе вещей должна осуществлять классификацию сигналов на две или на три группы (три группы возникают в том случае, если за счет наличия двух порогов выделяется еще и область, внутри которой отсутствует возможность определения изменений частоты). С этой точки зрения категориальный характер восприятия места образования согласных является не только объяснимым, но и логически необходимым.

Для определения положения порога, используемого испытуемыми, были вычислены отклонения среднего значения границы от частоты максимума гласного. Зависимость этих отклонений (вы-

Рис. 5. Зависимость среднего значения границы между [p] и [t] от частоты спектрального максимума гласного.

По оси абсцисс — частота спектрального максимума использовавшегося гласного, кгц; по оси ординат — среднее значение границы.

раженных в процентах) от частоты максимума гласного изображена на рис. 6, А. Можно видеть, что порог является частотно зависимым. Для того чтобы выяснить, не наблюдалась ли подобная зависимость в экспериментах по восприятию формантных переходов, была сделана дополнительная обработка данных, приведенных в работе Либермана и др. (Liberman et al., 1954). По объединенным данным четырех экспериментов (два значения перехода F_1 и идентификация глухих и звонких согласных) были построены кривые идентификации [p] (или [b]) для разных гласных (разных значений F_2 на стационарном участке).

В качестве границы была принята разность между начальным и стационарным значениями F_2 , соответствующая 50% распознавания [p] (или [b]). Величина этой разности, выраженная в процентах от F_2 , приведена на рис. 6, Б как функция от абсолютного

значения F_2 гласного. На этом же рисунке точками приведены данные 6 испытуемых, различавших [b], [d], [g], в новых экспериментах Либермана, докладывавшихся на Международном симпозиуме по речи в Киото (Liberman, 1968). Мы могли воспользоваться исходными графиками функций идентификации, любезно предоставленными автором. Данные чрезвычайно интересны тем, что показывают существенные различия в значениях порога у разных испытуемых.

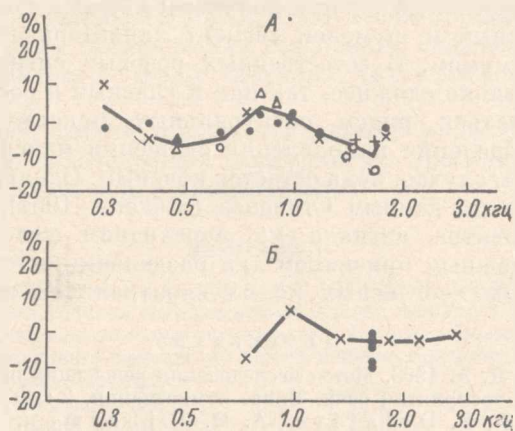


Рис. 6. Отклонения средних значений фонетических границ.

А — отклонение среднего значения границы между [p] и [t] от частоты максимума гласного. По оси абсцисс — частота максимума гласного, кгц; по оси ординат — отклонение среднего значения границы от частоты максимума гласного, %. Б — величина разности между значениями границы губные — язычные согласные и значениями второй форманты на стационарном участке гласного (по Liberman et al., 1954). По оси абсцисс — значения второй форманты, кгц; по оси ординат — величина разности, %.

Сравнение данных А и Б рис. 6 показывает, что в обоих случаях обнаруживается в принципе аналогичная частотная зависимость. В настоящее время трудно сказать, отражает ли она какие-то особенности слухового измерения направления изменения частоты максимума или же оказывается связанной с фонемной интерпретацией сигнала. Уменьшение отклонения границы от частоты максимума гласного (область несколько ниже 1000 гц на рис. 6, А, Б) означает, что порог устройства, индицирующего возрастание частоты и, таким образом, определяющего [p] (или [b]), понижается. Такого эффекта можно было бы ожидать, если бы в сигнале параллельно присутствовал еще какой-то признак,

повышающий вероятность выбора [p] (или [b]). Таким признаком могло бы быть абсолютное значение частоты шума или начальной частоты F_2 на формантном переходе.

ЗАКЛЮЧЕНИЕ

Полученные данные позволяют сделать вывод, что человек производит определение направления смещения спектрального максимума сигнала, и что он использует этот признак для разделения губных и язычных согласных.

Данные настоящей работы получены в упрощенных условиях, когда и гласный и шумовой сегмент характеризовались только одним максимумом. В естественных речевых сигналах дело обстоит значительно сложнее, так как и гласный и согласный могут характеризоваться рядом спектральных максимумов. Каковы правила определения направления смещения максимумов в этом более сложном случае, пока остается неясным. Однако существенно отметить, что по данным Стивенса (Stevens, 1967), направление смещения спектра сигнала на переходном участке является наиболее надежным признаком для разделения естественных губных и язычных согласных по их спектральной картине.

Л и т е р а т у р а

- Ч и с т о в и ч Л. А. 1968. Метод исследования решающих правил, применяемых при восприятии речи. *Вопр. языкознания*, 2 : 60—67.
- С о о п е р F. S., P. Delattre, A. M. Liberman, J. M. Borst, L. J. Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Amer.*, 24 : 597—606.
- Д е л а т т р е P. C., A. M. Liberman, F. S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Amer.*, 27 : 769—773.
- Ф л а н а г а н J. L. 1955. A difference limen for vowel formant frequency. *J. Acoust. Soc., Amer.*, 27 : 613—617.
- Н а р р и с K. S., H. S. Hoffman, A. L. Liberman, P. C. Delattre, F. S. Cooper. 1958. Effect of third-formant transitions on the perception of the voiced stop consonants. *J. Acoust. Soc. Amer.*, 30 : 122—126.
- Н о ф ф м а н H. S. 1958. Study of some cues in the perception of the voiced stop consonants. *J. Acoust. Soc. Amer.*, 30 : 1035—1041.
- Л и б е р м а н A. M. 1957. Some results of research on speech perception. *J. Acoust. Soc. Amer.*, 29 : 117—123.
- Л и б е р м а н A. M. 1968. Models of speech perception. *Speech Sympos., Kyoto*, D-2 : 1—4.
- Л и б е р м а н A. M., P. C. Delattre, F. S. Cooper, L. J. Gerstman. 1954. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.*, 68, 8 : 1—13.
- Л и б е р м а н A. M., F. S. Cooper, D. P. Shankweiler, M. Studert-Kennedy. 1967. Perception of the speech code. *Psychol. Rev.*, 74 : 431—461.
- П л о м п R. 1964. The ear as a frequency analyzer, I. *J. Acoust. Soc. Amer.*, 36 : 1628—1636.
- П л о м п R., A. M. Mimpfen. 1968. The ear as a frequency analyzer, II. *J. Acoust. Soc. Amer.*, 43 : 764—767.
- С т е в е н с K. N. 1967. Acoustic correlates of certain consonantal features. *Confer. on Speech Communication a. Processing, Boston* : 177—184.

ФОНЕТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ СТИМУЛОВ С РЕЗКИМИ ИЗМЕНЕНИЯМИ ИНТЕНСИВНОСТИ

Ю. И. Кузьмин, Д. М. Лисенко

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

Как показал ряд исследований, при распознавании звуков речи существенное значение имеют не столько абсолютные значения того или иного параметра сигнала, сколько характер их изменения во времени (Чистович, 1968; Кожевников и др., 1971). Вместе с тем роль динамических признаков, связанных с изменением интенсивности речевых сигналов, остается совершенно невыясненной.

В настоящей работе приводятся данные ряда экспериментов по восприятию речеподобных стимулов, на протяжении которых имелось резкое изменение интенсивности.

МЕТОДИКА

Стимулы синтезировались на установке, блок-схема которой приведена на рис. 1. В качестве источника сигнала использовался генератор сложных напряжений, позволявший независимо регулировать частоту основного тона и относительную интенсивность каждой из шести первых гармоник. Значения указанных параметров подбирались таким образом, чтобы полученный сигнал был близок по звучанию к русским гласным [а], [и] или [i]. С помощью двухканального электронного ключа формировались две посылки сигнала, длительность каждой из которых, как и длительность интервала между ними, задавалась при помощи специального блока регулировки длительности. Интенсивность посылок регулировалась при помощи аттенуатора. После суммирования этих сегментов получались сигналы, интенсивность которых изменялась, как это показано на рис. 2. Время нарастания интенсивности каждого из сегментов сигнала было в пределах 10 мсек.

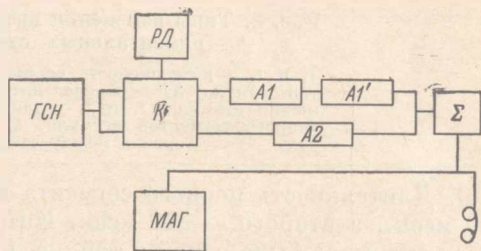


Рис. 1. Блок-схема установки для синтеза экспериментальных стимулов.

ГСН — генератор сложных напряжений; К — двухканальный электронный ключ; РД — блок регулировки длительности; А1, А1', А2 — аттенуаторы; Σ — сумматор; МАГ — магнитофон.

Предварительное прослушивание экспериментальных сигналов показало, что такого рода стимулы воспринимаются как слог, состоящий из согласного и гласного (СГ), если интенсивность второго сегмента превышает интенсивность первого на некоторую величину, а длительность первого сегмента остается в пределах 30—150 мсек. При больших длительностях первого сегмента сигнал воспринимается как трехзвучное сочетание типа ГСГ. Кроме того, было замечено, что качество согласного меняется в зависимости от величины перепада интенсивности.

Целью описываемых далее экспериментов было установить качественные и количественные зависимости, характеризующие это явление.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Первый эксперимент. Предполагалось определить набор согласных, воспринимаемых при прослушивании описанных выше сигналов, и установить характер зависимости между величиной перепада интенсивности и фонетической интерпретацией стимула. С этой целью был синтезирован набор стимулов составленных из двух отрезков гласного [а], следующих с нулевым интервалом. Основная частота сигнала была 200 гц. Интенсивности второй и последующих гармоник относительно интенсивности основного тона составляли соответственно 0,0, 0,5, 1,0, 0,5,

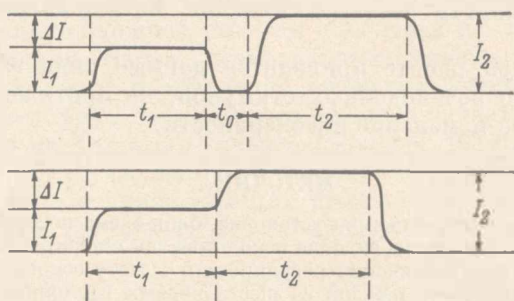


Рис. 2. Типы изменения интенсивности экспериментальных стимулов.

I_1 и I_2 — интенсивность первого и второго сегментов стимула; ΔI — перепад интенсивности на протяжении стимула; t_1 и t_2 — длительности первого и второго сегментов стимула, t_0 — длительность интервала между сегментами.

1,0. Длительность первого сегмента во всех случаях была равна 90 мсек., а второго — 220 мсек. Интенсивность второго сегмента всех стимулов была одной и той же, в то время как интенсивность первого менялась относительно интенсивности второго от 0 до —44 дб ступенями по 4 дб. Таким образом, всего было получено 12 разных стимулов. Из этих стимулов была составлена и записана на магнитофонную ленту таблица, в которой каждый стимул встречался в случайном порядке 10 раз.

Таблица была дважды предъявлена группе слушателей из 11 человек через громкоговорители в звукозаглушенной комнате. Уровень воспроизведения составлял около 60 дб над порогом восприятия. От испытуемых требовалось как можно точнее записать каждый стимул, пользуясь буквами русского алфавита (набор возможных ответов не ограничивался). Всего на каждый стимул было получено по 220 суждений. Полученные результаты приведены в табл. 1.

Хотя набор сочетаний, использованных слушателями, достаточно разнообразен, данные сравнительно легко систематизи-

Процентные соотношения разных сочетаний, использованных слушателями
при записи стимулов в первом эксперименте

№ сти- мула	ΔI , дБ	Использованные сочетания																
		$a(n)$ $a(m)$	аа	$na(n)$ $ta(n)$	$naa(n)$	$aa(n)$	$ma(n)$	$maa(n)$	aaa $и аа$	$тааа(n)$ $кла$	маа	и а $у а$	$ба(н)$ $да(н)$ $ваб$	$баа(н)$	на	$и та$ $ата$	$аб$ $ам$ $там$	ама
1	0	81.3	4.1	3.6	7.7	0.9	—	—	—	—	—	—	0.5	0.5	0.9	—	0.5	—
2	4	36.4	5.5	8.6	7.7	15.5	2.3	—	11.4	0.9	7.2	0.9	—	0.5	3.2	—	—	—
3	8	10.0	1.8	4.1	0.5	20.5	15.0	4.1	16.4	5.5	3.6	5.5	0.5	—	11.8	0.9	—	—
4	12	3.6	0.5	0.5	—	15.0	26.8	4.1	15.9	7.2	2.7	6.8	1.4	—	13.6	0.9	0.9	—
5	16	5.5	0.5	0.5	—	14.5	30.5	4.5	9.1	7.7	2.7	5.9	0.9	—	16.4	0.9	0.9	0.5
6	20	5.0	0.5	1.8	—	13.2	36.4	2.7	4.5	6.4	4.5	5.0	—	1.4	17.3	1.4	—	—
7	24	10.0	0.9	0.9	—	15.4	40.0	1.4	3.2	6.8	0.9	3.2	—	—	16.8	—	0.5	—
8	28	18.2	—	3.2	—	14.1	37.3	0.9	1.8	2.7	1.8	2.3	0.5	1.4	15.9	—	—	—
9	32	34.5	—	2.7	—	12.3	33.6	0.5	0.5	0.5	0.5	0.5	0.5	1.8	11.4	—	0.9	—
10	36	51.8	—	5.5	1.4	5.9	15.0	0.9	—	0.9	—	—	0.4	8.2	9.1	—	0.9	—
11	40	66.3	—	10.5	0.5	2.3	33.6	—	—	—	—	—	—	10.9	5.5	—	0.5	—
12	44	74.0	—	15.4	0.5	—	0.5	—	—	—	—	—	—	5.9	2.7	—	0.9	—

руются. Как видно из этой таблицы, при $\Delta I \leq 4$ дб слушатели записывают стимулы как отдельные гласные или как гласные, которым предшествует глухой смычный согласный (как правило, *п*). При $\Delta I > 4$ дб стимулы интерпретируются как слоги, содержащие согласные *л*, *м*, *н*. При $\Delta I \geq 36$ дб стимулы опять интерпретируются главным образом как *а* или *па*. Кроме того, одним из слушателей такие стимулы записывались как слог *ба*.

Можно видеть, что буквы, обозначающие глухие смычные согласные, появляются в ответах испытуемых при любых значениях ΔI . Более того, процент таких ответов распределен по

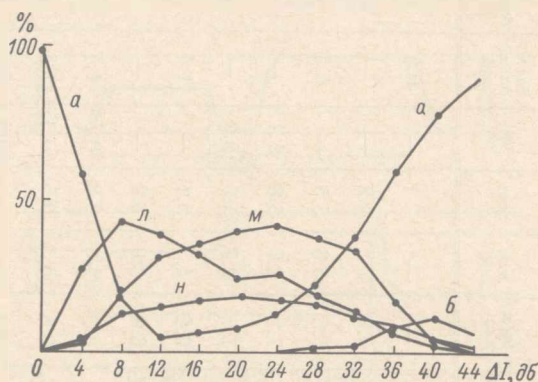


Рис. 3. Распределение ответов при записи стимулов с разными перепадами интенсивности.

По оси абсцисс — перепад интенсивности, дб; по оси ординат — процент разных ответов.

шкале ΔI приблизительно равномерно. Следовательно, основным эффектом изменения интенсивности сигнала следует считать появление в ответах испытуемых *л*, *м*, *н* и, возможно, *б*.

На рис. 3 показана частота восприятия разных слогов, содержащих указанные согласные, при разных значениях ΔI . При построении данного графика не были учтены слоги типа *мна* и *мла*, содержащие по два интересующих нас согласных, и слоги типа *иа*, *ам* и т. п. Общее количество таких ответов составило 5.6%. Из графика видно, что частота восприятия разных согласных зависит от величины перепада интенсивности. При меньших значениях ΔI записывается преимущественно *л*, при больших — *м* или *н*. При ΔI порядка 40 дб и более может появляться *б* (по данным одного испытуемого). При этом слоги, содержащие *м* и *н*, распределены в одном и том же интервале изменений ΔI .

Второй эксперимент. С целью получить предварительные данные о влиянии абсолютного уровня сигнала на способ его фонетической интерпретации та же таблица стимулов была дважды предъявлена группе из 10 слушателей при двух

разных уровнях воспроизведения. В данном случае испытуемым предлагалось записывать стимулы только одним из четырех дозволённых сочетаний *a*, *ла*, *ма* и *на*. В конце каждого воспроизведения таблицы определялся порог восприятия синтетического [а] длительностью 90 мсек. Уровень воспроизведения таблицы определялся относительно этого порога.

Полученные результаты приведены на рис. 4. Можно видеть, что в обоих случаях распределение ответов остается практически одним и тем же. Это позволяет думать, что фонетическая интерпретация исследуемых стимулов зависит только от величины перепада интенсивности и существенно не зависит от его абсолют-

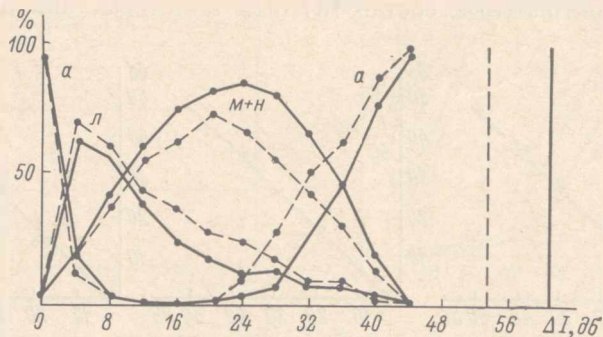


Рис. 4. Распределение ответов испытуемых при двух разных уровнях интенсивности воспроизведения стимулов.

По оси абсцисс — перепад интенсивности, дБ; по оси ординат — процент разных ответов. Вертикальными линиями обозначены интенсивности второго сегмента стимула относительно порога восприятия изолированного гласного, штриховая линия соответствует более низкому уровню воспроизведения.

ного уровня. Дополнительные данные, говорящие в пользу этого предположения, были получены в следующем эксперименте.

Третий эксперимент. Его основная задача заключалась в том, чтобы установить, является ли величина перепада интенсивности, необходимая для образования разных согласных, одной и той же вне зависимости от спектрального состава стимула. Судя по предварительным прослушиваниям, набор согласных и порядок их появления по мере изменения перепада интенсивности остаются одними и теми же при любом спектральном составе стимула. Поэтому было решено, не получая полных распределений типа приведенных на рис. 4, определить для стимулов с разными спектрами только те критические величины ΔI , которые соответствуют границам между разными слогами.

В эксперименте была использована методика активного поиска границ (Чистович, 1968). Сущность процедуры заключалась в сле-

дующем. Испытуемый, оперируя аттенуатором I' (рис. 1), должен был подобрать такую интенсивность первого сегмента стимула, при которой один слог начинал переходить в другой. Перед началом каждого нового измерения экспериментатор устанавливал некоторое значение на аттенуаторе I , так что исходная интенсивность первого сегмента варьировала в определенных пределах.

Таким способом были определены две границы: граница, соответствующая переходу изолированного гласного в слог с согласным [l], и граница, соответствующая переходу слога с согласным [l] в слог, содержащий [m] или [n]. Обе границы были определены для каждого из трех синтетических гласных [a], [u] и [i]. Спектральный состав [a] был таким же, как и в первом

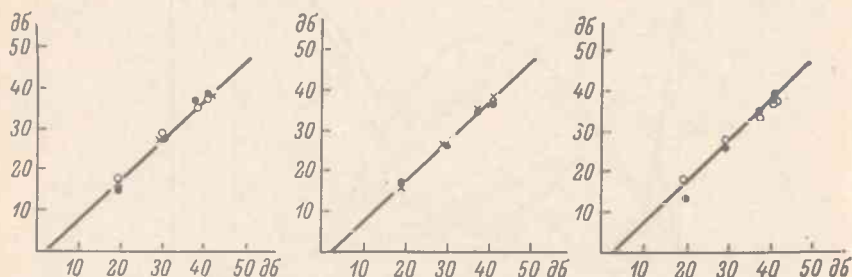


Рис. 5. Зависимость интенсивности первого сегмента стимула (ось ординат) от уровня интенсивности второго.

Левый график — для гласного [a], средний — [u], правый — [i].

эксперименте. Относительная интенсивность второй и последующих гармоник в спектрах [u] и [i] была равна 1.0, 1.0, 0.0, 0.5, 0.0 и 0.0, 0.0, 0.3, 1.0, 1.0 соответственно. Частота основного тона [u] была равна 200 гц, а частота основного тона [i] — 400 гц.

В эксперименте по определению первой границы участвовало трое испытуемых. Каждый из них определял границу от 5 до 10 раз для каждого из четырех разных значений интенсивности второго сегмента стимула. Полученные результаты приведены на рис. 5. Можно видеть, что при всех значениях интенсивности второго сегмента переход гласного в слог происходит приблизительно при одном и том же значении ΔI (зависимость хорошо аппроксимируется прямой с наклоном 45°). Это вполне согласуется с выводом, сделанным на основании предыдущего эксперимента. Кроме того, величина ΔI остается приблизительно одной и той же для всех исследованных гласных и составляет 2.5—3.5 дБ. Это также согласуется с данными предыдущего эксперимента, судя по которым перепад интенсивности, при котором стимул с равной вероятностью записывается испытуемым как a и la , приблизительно равен 2.5 дБ (рис. 4). Таким образом, можно считать, что стимулы начинают восприниматься как слоги при одном и том же перепаде

интенсивности независимо от их спектрального состава и от абсолютного уровня интенсивности.

Учитывая этот результат, границы, соответствующие переходу слогов [la], [lu], [l'i] в слоги, содержащие назальные согласные, определялись только для одного значения интенсивности второго сегмента. В определении этих границ участвовало четверо испытуемых. Полученные результаты приведены в правой части рис. 6. По горизонтали отложена величина перепада интенсивности ΔI , значками указаны определенные разными испытуемыми положе-

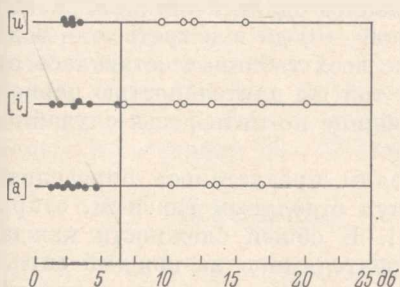


Рис. 6. Положение границ, соответствующих переходу отдельных гласных в слоги, содержащие [l] (точки) и переходу слогов, содержащих [l], в слоги с назальными согласными (кружки).

По горизонтали — величина перепада интенсивности, дБ; вертикальные линии — значения ΔI приблизительно соответствующие пикам распределений, приведенных на рис. 4.

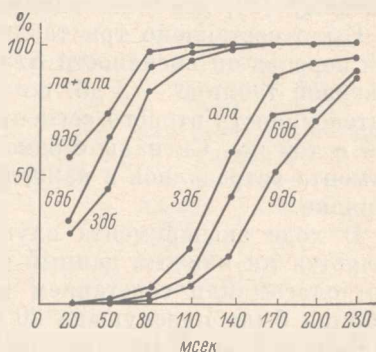


Рис. 7. Зависимость интерпретации стимула от длительности первого сегмента.

По оси абсцисс — длительность первого сегмента стимула, мсек.; по оси ординат — процент записи стимула в виде сочетания, содержащего согласный (левые кривые) и в виде сочетания ала (правые кривые). Параметр кривых — величина перепада интенсивности.

ния границ. В левой части того же рисунка отмечены границы, полученные из графиков рис. 5. Вертикальные линии соответствуют максимумам распределений, полученных во втором эксперименте (рис. 4). Можно видеть, что положение второй границы значительно флюктуирует как от гласного к гласному, так и от испытуемого к испытуемому. Однако, как было обнаружено во втором эксперименте, значительной флюктуации этой границы сопутствует сильное переслоение распределений, характеризующих частоту восприятия соответствующих звуков. В то же время какой-либо зависимости между положением этой границы и изменением спектрального состава стимула не обнаруживается.

Четвертый эксперимент. Все предыдущие эксперименты были проведены на стимулах, длительность первого сегмента которых была равна 90 мсек. В данном эксперименте предполагалось проверить, зависит ли величина перепада интен-

сивности, необходимая для превращения гласного в слог, от длительности первого сегмента. Одновременно предполагалось определить ту длительность первого сегмента, при которой он начинает восприниматься как гласный.

С этой целью на той же установке были синтезированы сигналы, длительность первого сегмента которых варьировала от 20 до 230 мсек. ступенями по 30 мсек. Длительность второго сегмента стимула оставалась во всех случаях одной и той же и равной 220 мсек. Спектральные характеристики сигналов соответствовали гласному [a], использованному в предыдущих экспериментах.

Было составлено три таблицы таких сигналов. В каждой из них перепад интенсивности ΔI оставался одним и тем же и равным в первой таблице — 3 дБ, во второй — 6 дБ и в третьей — 9 дБ. Интенсивность второго сегмента во всех таблицах оставалась одной и той же. Сигналы с одной и той же длительностью первого сегмента встречались в каждой таблице по пять раз в случайном порядке.

В ходе эксперимента слушателям предлагалось определить, является ли каждый данный стимул отдельным гласным, открытым слогом или сочетанием [ala]. В общей сложности каждая таблица была прослушана 20 слушателями и на каждый сигнал было получено по 100 оценок. Полученные результаты приведены на рис. 7.

Из графика следует, что при всех перепадах интенсивности процент восприятия согласного увеличивается по мере увеличения длительности первого сегмента.

Ответ на второй из интересовавших нас вопросов дают правые кривые графика. Можно видеть, что при длительности первого сегмента менее 50 мсек. стимул вообще не может быть воспринят как сочетание ГСГ. При ΔI , равном 3 дБ, процент восприятия таких сочетаний, равный 75, достигается при длительности первого сегмента порядка 150 мсек. При больших перепадах интенсивности та же величина восприятия сочетания ГСГ достигается при больших длительностях первого сегмента.

При интерпретации последнего факта необходимо иметь в виду следующее обстоятельство. Поскольку интенсивность второго сегмента стимула во всех случаях оставалась одной и той же, в таблицах с большими значениями перепада интенсивности интенсивность первого сегмента была меньше.

Пятый эксперимент. В ходе предварительных экспериментов было замечено, что стимул, состоящий из двух отрезков гласного, может восприниматься как слог типа [la] или сочетание типа [ala] также и в том случае, если между отдельными сегментами стимула имеется небольшая пауза. В данном эксперименте предполагалось определить максимальную длительность этой паузы, при которой это явление еще наблюдается. На этот раз экспериментальные стимулы синтезировались из отрезков

двух гласных — [u] и [a]. Длительность первого из них варьировала от 66 до 260 мсек., а длительность второго оставалась во всех случаях одной и той же и равной 184 мсек. В ходе эксперимента испытуемые должны были подобрать ту минимальную длительность интервала между гласными, при которой они начинают восприниматься без призвуча согласного [1].

В эксперименте участвовало трое испытуемых, каждый из которых определял указанный интервал по 10 раз для каждой из пяти разных длительностей первого сегмента. Полученные результаты приведены в табл. 2. Можно видеть, что длительность определяемого интервала остается примерно одинаковой при всех длительностях первого сегмента. Кроме того, средние длительности интервала оказались довольно близкими у разных испытуемых и составили 32—53 мсек.

Т а б л и ц а 2

Средние длительности интервала между сигналами (мсек.) при разных длительностях первого сегмента (мсек.)

Длительность первого сегмента	Испытуемые		
	1-й	2-й	3-й
66	30	55	41
115	33	59	49
163	35	45	53
212	33	53	52
260	28	54	52
Среднее	32	53	50

ОБСУЖДЕНИЕ

Как показывает весь комплекс полученных данных, резкое изменение интенсивности сигналов во времени является существенным признаком при их фонетической интерпретации. Изменяя величину перепада интенсивности на протяжении гласного, можно получить стимулы, воспринимаемые как слоги, содержащие разные согласные. Указанный эффект не зависит от спектра стимула и от абсолютного уровня его интенсивности.

Для объяснения этих фактов необходимо, однако, уточнить, действительно ли человек, проводя классификацию исследуемых стимулов, основывается на результатах измерения интенсивности. Как известно, резкое изменение интенсивности сигнала неизбежно связано с изменением его спектра. Таким образом, можно предположить, что возникновение согласного при восприятии таких сигналов связано именно с этим обстоятельством. Чтобы проверить это предположение, были проведены эксперименты, в которых использовались сигналы, составленные из отрезков гласного одинаковой интенсивности с небольшим интервалом между ними, а также сигналы, состоящие из синтетического гласного с наложенным на него отрезком того же гласного длительностью 5 мсек.

Изменения спектра в этих сигналах, обусловленные быстрым изменением интенсивности, должны быть аналогичны тем, которые имели место в стимулах, использованных в основных экспериментах. Тем не менее такие стимулы не воспринимались как

слоги. Это позволяет думать, что при интерпретации стимулов, исследованных в данной работе, человек использует в качестве полезного признака именно изменение интенсивности как таковое. В пользу этого вывода говорит также следующее обстоятельство.

Если бы появление согласного было обусловлено только изменением спектра на протяжении переходного процесса, вероятность восприятия согласного не должна была бы зависеть от длительности первого сегмента стимула, если только эта длительность превышает время нарастания сигнала плюс время установления процесса в улитке. Так как время нарастания использованных нами сигналов оставалось в пределах 10 мсек., а время установления процесса в улитке при частоте основного тона 200 гц не превышает 10 мсек. (Békésy, 1960; Молчанов, Лабутий, 1967), вероятность восприятия согласного должна бы оставаться одной и той же для всех стимулов, длительность первого сегмента которых более 20 мсек. На самом деле, как показывают данные четвертого эксперимента, вероятность восприятия согласного увеличивается по мере увеличения длительности первого сегмента вплоть до величины порядка 140 мсек. Кроме того, как было показано в заключительном эксперименте, согласный воспринимается также и в том случае, если между двумя посылками с разной интенсивностью имеется небольшая пауза. Таким образом, оба эти факта позволяют думать, что условием, необходимым для появления согласного, является скорее наличие разницы в интенсивности стимула, чем связанные с изменением интенсивности изменения спектра.

Л и т е р а т у р а

- Кожевников В. А., Л. Е. Рейтблат, Л. А. Чистович. 1971. Знак изменения частоты спектрального максимума как различительный признак места образования согласного. Наст. сб. : 74—82.
Молчанов А. П., В. К. Лабутий. 1967. Слух и анализ сигналов. Л.
Чистович Л. А. 1968. Метод исследования решающих правил, применяемых при восприятии речи. Вопр. языкознания, 2, 60—67.
Békésy G. V. 1960. Experiments in hearing. N. Y.

ВОСПРИЯТИЕ ДЛИТЕЛЬНОСТИ ЗВОНКИХ СОГЛАСНЫХ

М. Г. Жукова

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

Исследование восприятия длительности отдельных отрезков речевого сигнала представляет большой интерес, так как измерение длительности логически предполагает предварительное выделение начала и конца измеряемого явления. Производит ли нервная система предварительную сегментацию речевого потока,

на какие сегменты производится деление, что принимается за начало и конец измеряемого явления — все эти вопросы пока остаются неясными.

Целью настоящей работы являлось исследование восприятия и воспроизведения длительности звонкого участка смычки русских начальных согласных [b], [d], [m] в слогах типа СГ. Длительность согласного должна учитываться при восприятии и может быть интерпретирована по-разному. Во-первых, можно предположить, что должна существовать минимальная пороговая длительность, ниже которой согласный пропускается или может быть воспринят как другой согласный. При длительности же больше пороговой она может служить для различения одиночных и двойных согласных и может, вероятно, учитываться при определении темпа.

Работа разделяется на две части. Первая серия экспериментов была проделана для того, чтобы выяснить, действительно ли существует пороговая длительность звонкого согласного. Во второй части работы исследовалось восприятие и воспроизведение надпороговой длительности звонкого согласного.

ОПРЕДЕЛЕНИЕ ПОРОГОВОЙ ДЛИТЕЛЬНОСТИ ЗВОНКОГО СОГЛАСНОГО

Исследование пороговой длительности звонкого согласного было начато с проведения экспериментов с естественными слогами [bu] и [du], у которых отрезался начальный участок звонкой смычки согласного. Опыты по фонемной идентификации этих сигналов показали, что укорочение длительности звонкого участка согласного до 10—15 мсек. превращает звонкий согласный в соответствующий ему глухой взрывной согласный. Так как имеются данные о том, что при укорочении длительности щелевых согласных до величины, не превышающей 20—25 мсек., они также превращаются в глухие взрывные согласные [p] и [t] (Деркач, 1958), возможно предположить, что превращение звонкого согласного в глухой связано с явлением, не имеющим прямого отношения к собственно измерению длительности согласного, а именно — с утратой сигналом тонального характера и превращением его в щелчок. Если это предположение правильно, то любой звонкий согласный, в том числе и сонант, при сокращении его длительности должен превращаться в глухой согласный.

Для проверки этого предположения был проведен эксперимент по определению пороговой длительности согласного [m] в синтетическом слоге [ma], полученном с помощью установки, блок-схема которой приведена на рис. 1.

Измерения производились на трех частотах основного тона сигнала (200, 300 и 400 гц), который в каждом случае был одним и тем же для сегментов, соответствующих согласному и гласному. Отношения амплитуд высших гармоник

к амплитуде основного тона сигнала для сегментов, соответствующих согласному и гласному, при различных значениях частоты основного тона сигнала, приводятся в таблице.

Отношение амплитуд высших гармоник к амплитуде
основного тона сигнала

Основной тон, Гц	Номера гармоник									
	согласный					гласный				
	2	3	4	5	6	2	3	4	5	6
200	0.2	—	0.6	0.3	—	0.5	0.5	1.0	0.5	1.0
300	—	—	0.3	—	—	0.5	1.0	1.0	—	—
400	—	0.2	—	—	—	0.5	1.0	1.0	—	—

Для определения значения пороговой длительности согласного был применен метод активного поиска границ. Испытуемым была предоставлена возможность самостоятельно изменять длительность сегмента, соответствующего согласному, их задача заключалась в том, чтобы подобрать такое значение его длительности, при котором звонкое [m] начинает звучать как глухое [p].

При каждом значении частоты основного тона сигнала было произведено по 100 определений границы (по 10 определений для каждого из 10 испытуемых).

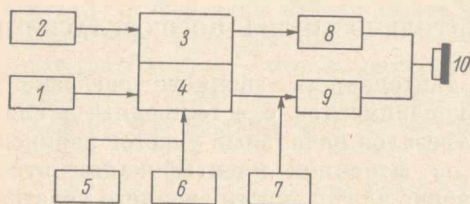


Рис. 1. Блок-схема установки для получения синтетического слога.

1 — генератор сложного напряжения ГСН-1 (гласный); 2 — генератор сложного напряжения ГСН-1 (согласный); 3, 4 — двухканальный электронный ключ; 5 — магазин емкостей для регулирования длительности согласного экспериментатором; 6 — магазин емкостей для регулирования длительности согласного испытуемым; 7 — измеритель времени; 8, 9 — магазины затухания; 10 — электродинамический телефон ТД-6.

длительность звонкого согласного, которая определяется человеком с большой точностью, и ее величина равна в среднем 20 мсек. Поскольку сонант [m] имеет периодический характер, а глухое [p] не является сигналом периодическим, естественно считать, что найденный порог (20 мсек.) отражает минимальную длительность сигнала, необходимую для обнаружения его периодического характера. В пользу этого предположения говорят также данные Рисберга (Risberg, 1968) о том, что человек перестает различать периодический (сложный сигнал

На основании полученных данных были вычислены средние значения пороговой длительности звонкого согласного и величины ее среднего квадратического отклонения. Как показывает рис. 2, эти значения мало зависят от частоты основного тона сигнала, а сама величина среднего квадратического отклонения является незначительной. Следовательно, действительно имеется пороговая

с основной частотой 200 гц) и шумовой характер сигнала при их длительности, равной 10—20 мсек., если спектр шумового сигнала

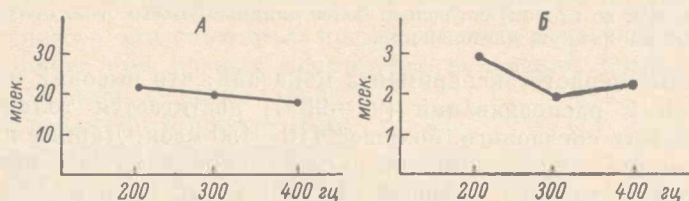


Рис. 2. Зависимость пороговой длительности звонкого согласного (А) и ее среднего квадратического отклонения (Б) (по оси ординат, мсек.) от частоты основного тона сигнала (по оси абсцисс, гц).

находится в низкочастотной области (полоса 300—800 гц), и при длительности 10 мсек., если спектр шумового сигнала находится в высокочастотной области (полоса 2400—3400 гц).

ВОСПРИЯТИЕ НАДПороговой длительности звонкого согласного

Исследование восприятия надпороговой длительности звонкого согласного было начато с проведения экспериментов по парному сравнению субъективной длительности согласных.

В первом, предварительном, эксперименте использовались естественные слоги [bu] и [du], у которых отрезался начальный участок звонкой смычки согласных. Слоги были объединены в пары так, что длительность зонкого участка согласного в одном из слогов (в первом или во втором) была всегда одной и той же (15—20 мсек.), в другом же слоге она была различной в разных парах. Максимальная длительность согласного в слоге [bu] была равна 190 мсек. и в слоге [du]—250 мсек. Задача испытуемого заключалась в том, чтобы при прослушивании сигналов определить, в каком слоге согласный более длинный.

Во втором эксперименте использовался тот же синтетический слог [ma], что и в опытах по определению пороговой длительности согласного (основной тон 300 гц). Длительность сегмента, соответствующего гласному, всегда оставалась равной 500 мсек., длительность же сегмента, соответствующего согласному, изменялась в пределах от 25 до 205 мсек. ступенями по 10 мсек. Из полученных исходных слогов была составлена таблица, в которой слоги были объединены в пары. В каждую пару входил слог с постоянной длительностью согласного 25 мсек. и слог, длительность которого была различной в разных парах.

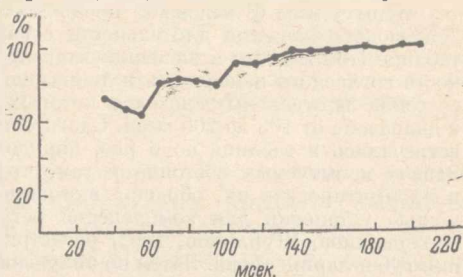


Рис. 3. Процент правильных распознаваний длительности звонкого согласного (по оси ординат) в зависимости от длительности предъявляемого согласного (по оси абсцисс, мсек.).

Слоги с постоянной длительностью согласного встречались в таблице одинаковое число раз на первом и на втором месте в паре. Слоги с одним и тем же значением переменной длительности согласного встречались в таблице по 6 раз. Испытуемые при прослушивании этой таблицы определяли, в каком слоге (в первом или во втором) согласный более длинный. Всего было получено по 69 ответов на каждую длительность.

Данные первого эксперимента показали, что высокий процент правильных распознаваний (90—95%) достигается только при длительности согласного, большей 110—150 мсек. Данные второго эксперимента, приведенные на рис. 3, показывают, что при длительности согласного, меньшей 100—110 мсек., процент, правильных распознаваний длительности колеблется от 60 до 80% и что четкое различие наступает при длительности согласного, большей 120 мсек.

Таким образом, на основании этих данных можно было предполагать, что в пределах до 100—120 мсек. нет четкого различия длительности согласного.

ВОСПРОИЗВЕДЕНИЕ НАДПороГОВОЙ ДЛИТЕЛЬНОСТИ ЗВОНКОГО СОГЛАСНОГО

Для исследования воспроизведения длительности звонкого согласного использовался метод имитации. Одно из преимуществ этого метода состоит в том, что он позволяет достаточно точно выяснить, длительность какого именно элемента слога воспринимает и имитирует испытуемый, в то время как в психоакустических экспериментах это сделать невозможно. В экспериментах по имитации использовался синтетический слог [ma] (основной тон 300 гц).

Испытуемым (3 человека) предъявлялось две таблицы слогов, в которых диапазоны изменений длительности согласного были различными. В первой таблице (той же, что и в экспериментах по субъективному сравнению) длительность согласного изменялась в диапазоне от 25 до 205 мсек. Вторая таблица состояла из отдельных слогов, в которых длительность согласного изменялась в диапазоне от 105 до 205 мсек. Слоги с одинаковой длительностью согласного встречались в таблице по 6 раз, порядок их предъявления был случайным. Задача испытуемых состояла в том, чтобы прослушать полностью сигналы и проимитировать их, обращая внимание на длительность согласного. С помощью установок для комплексной регистрации артикуляторных движений (Кожевников, Шупляков, 1962) регистрировались показания датчика смыкания губ и ларингофона. Затем по полученным осциллограммам измерялась длительность звонкого согласного и всего слога. На каждую предъявленную длительность как для первой, так и для второй таблицы было получено от 60 до 70 ответов.

Из рис. 4 видно, что при увеличении длительности предъявляемого согласного общая длительность слога значительно увеличивается, в то время как длительность гласного остается почти без изменений. Уже на этом основании можно утверждать, что испытуемые определяют момент перехода от согласного к гласному и имитируют длительность согласного звука. Рис. 5 и 6 показывают

зависимость длительности имитирующего согласного и величины ее среднего квадратического отклонения от длительности предъявляемого согласного. На основании этих данных можно предположить, что зависимость между длительностью имитирующего согласного и длительностью предъявляемого согласного является монотонной при разных диапазонах изменений предъявляемых длительностей и носит аналоговый характер. В пользу этого предположения говорят следующие факты. В обоих диапазонах изменений предъявляемых длительностей согласного точки достаточно хорошо аппроксимируются прямыми за исключением одного случая (испытываемый В.). Величина же среднего квадратического отклонения хотя и несколько увеличивается при увеличении длительности предъявляемого согласного, но не превышает в среднем 35—40 мсек. и не имеет резких пиков на каких-либо отдельных участках. Однако по абсолютным значениям длительность имитирующего согласного превышает длительность предъявляемого согласного, и наклон аппроксимирующих прямых различен как у разных испытуемых в пределах одного и того же диапазона, так и у одного и того же испытуемого при разных диапазонах изменений длительности предъявляемого согласного. Кроме того, угловые коэффициенты наклона аппроксимирующих прямых, вычисленные методом наименьших квадратов, в случае более узкого диапазона изменений предъявляемых длительностей согласного у всех испытуемых несколько больше (1.07, 0.97, 0.93), чем в случае более широкого диапазона (0.66, 0.86, 0.83), но в данном случае это различие не является статистически достоверным и требует дополнительной проверки. Следует отметить, что аналогичные зависимости между входными и выходными сигналами, а также факт зависимости наклона аппроксимирующих прямых от величины диапазона изменений предъявляемых параметров были получены при имитации громкости (Malinnikova, 1968) и высоты сигналов (Люблинская, 1968).

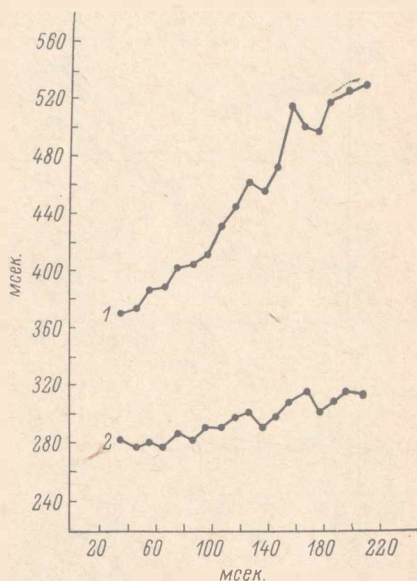


Рис. 4. Изменение длительности всего слога (1) и длительности гласного (2) (по оси ординат, мсек.) в зависимости от длительности предъявляемого согласного (по оси абсцисс, мсек.) при имитации.

ОБСУЖДЕНИЕ

В работе Абрамсона и Лискера (Abramson, Lisker, 1967) указывается, что временные отношения между началом голосового возбуждения и взрывом согласного определяют различие на-

чальных глухих и звонких согласных [p]—[b], [t]—[d], [g]—[k] и что четкое различие изменений временного промежутка между этими явлениями наблюдается только на фонемных границах. На основании этих результатов был сделан вывод о том, что восприятие длительности звонких согласных носит категориальный характер, а длительность в пределах одной категории не различается.

В настоящей работе также были получены данные о том,

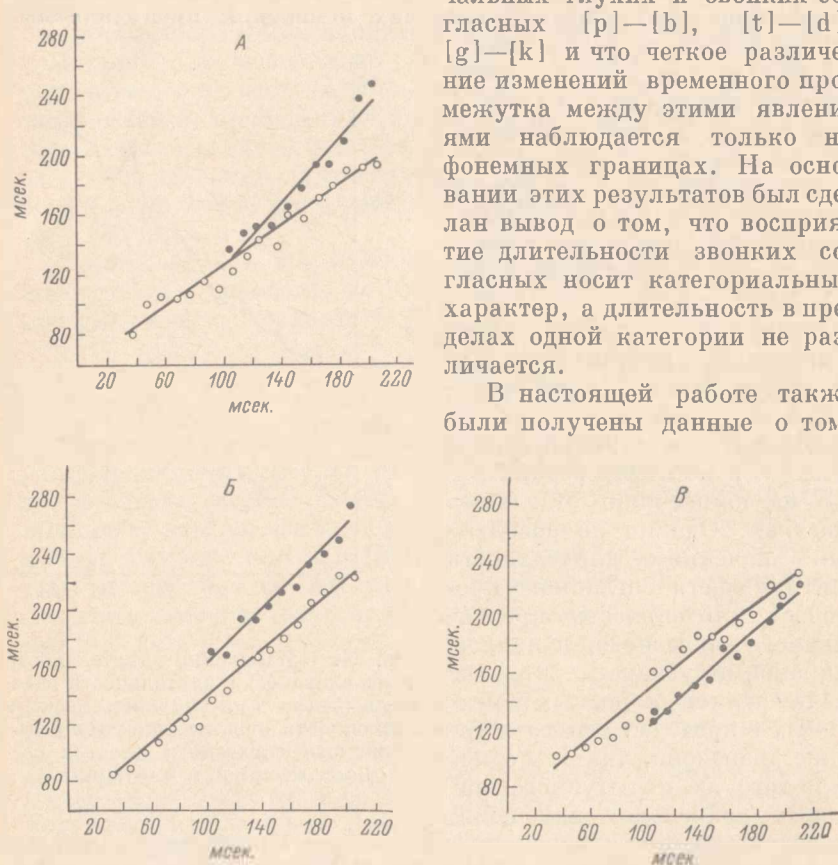


Рис. 5. Зависимость длительности имитирующего согласного (по оси ординат, мсек.) от длительности предъявляемого согласного (по оси абсцисс, мсек.) при двух различных диапазонах изменения длительности предъявляемого согласного.

А, Б, В — разные испытуемые.

что существует пороговая длительность звонких согласных, которая, однако, должна рассматриваться как порог различения периодической структуры сигнала. С этой точки зрения восприятие звонких и глухих согласных действительно может быть категориальным.

Рассмотрим далее данные о восприятии и воспроизведении надпороговой длительности звонкого согласного. На основании данных, полученных в экспериментах по субъективному сравнению длительности согласных, и данных Абрамсона и Лискера действительно можно было бы утверждать, что человек недостаточно четко различает длительность согласного, по крайней мере при ее величине, меньшей 100—120 мсек. Однако данные, полученные в опытах по имитации, не подтверждают этого предположения и позволяют утверждать, что испытуемый имитирует, а следовательно, и различает длительность звонкого согласного. Так как зависимость между длительностью имитирующего и предъявляемого согласных сохраняет монотонный характер при обоих диапазонах изменений длительности предъявляемого согласного, можно считать, что различение, а следовательно, и измерение длительности начинается с величин, очень близких к пороговому значению длительности звонкого согласного.

Измерение длительности логически предполагает выделение начала и конца измеряемого явления. Однако на основании полученных данных трудно судить о точности, с которой производится выделение длительности согласного в последовательности СГ. Дело в том, что средняя квадратическая ошибка имитации складывается из ошибок: а) членения, б) измерения длительности входного сигнала и преобразования ее в соответствующий сигнал управления при речеобразовании, в) запоминания сигнала управления, г) реализации движений. Примем, что все ошибки, кроме ошибки членения, относительно малы и ими можно пренебречь. Примем также, что точность определения момента начала сегмента согласного равна точности определения момента перехода от согласного к гласному. Тогда из полученных данных следует, что средняя квадратическая ошибка членения составляет примерно 20 мсек. Очевидно, что это максимально возможная ошибка, т. е.

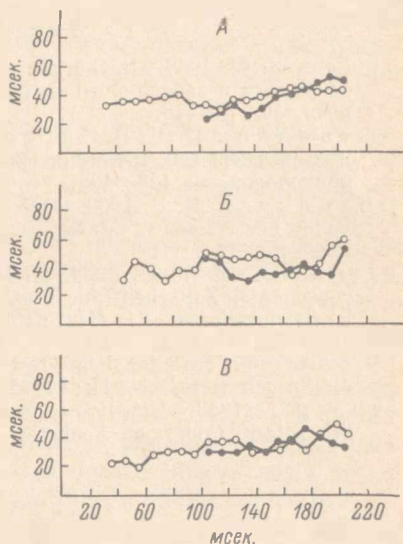


Рис. 6. Зависимость величины среднего квадратического отклонения длительности имитирующего согласного (по оси ординат, мсек.) от длительности предъявляемого согласного (по оси абсцисс, мсек.) при двух различных диапазонах изменения длительности предъявляемого согласного.

А, Б, В — разные испытуемые.

можно утверждать, что действительная средняя квадратическая ошибка членения не превышает в данном случае 20 мсек.

Таким образом, данные, полученные в настоящей работе, подтверждают предположение о том, что надпороговая длительность звонкого согласного различается с большой точностью и может учитываться при восприятии.

Л и т е р а т у р а

- Д е р к а ч М. Ф. 1958. Статистика восприятия глухих взрывных и щелевых согласных в зависимости от их длительности. В сб.: Вопр. статистики речи, Л.: 40—45.
- К о ж е в н и к о в В. А., В. С. Ш у п л я к о в. 1962. О приемах одновременной регистрации артикуляторных и акустических параметров. Вопр. психологии, 6, 128—134.
- Л ю б л и н с к а я В. В. 1968. Особенности имитации высоты тональных сигналов испытуемыми без «музыкального» слуха. Тр. VI Всесоюзн. акуст. конфер., М. 3 III-8.
- M a l i n n i k o v a T. G. 1968. Mimicking of loudness of synthetic fricative consonants. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 21 (1/2): 135—139.
- A b r a m s o n A. S., L. L i s k e r. 1967. Discriminability along the voicing continuum: cross-language tests. Haskins Laboratories Status Reports on Speech Research, 11: 17—22.
- R i s b e r g A. 1968. Periodic-nonperiodic test of hearing capacity. Reports of 6th Intern. Congr. on Acoustics, Tokyo, B: 219—222.

О ФУНКЦИОНАЛЬНОЙ МОДЕЛИ МЕХАНИЗМА СЕГМЕНТАЦИИ ТЕКУЩЕГО ПОТОКА СИГНАЛОВ

А. П. Молчанов

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

При изучении характера ответов одиночных нейронов в различных отделах сенсорных систем бросается в глаза обширная группа элементов, реагирующих только на начало или конец стимула.

С точки зрения функциональной модели восприятия сигналов нейроны, выделяющие моменты резкого изменения сигналов, могут рассматриваться как часть специального механизма сегментации, разделяющего непрерывный текущий поток сигналов на участки (сегменты), отличающиеся относительным постоянством физических характеристик.

Такой гипотетический механизм может выделять из потока сигналов признак, характеризующий его временную структуру независимо от типа сегментов, несущих информацию о смысловом содержании звуков.

Характерно, что задача сегментации, как правило, возникает и на определенном этапе решения технической проблемы автоматического распознавания звуков, в частности речевых.

Сегментация текущего речевого потока на основе объективных акустических параметров в настоящее время привлекает внимание целого ряда исследователей. Опубликованные работы показали, что, выделяя в речевом потоке моменты времени, в которые физические характеристики сигнала претерпевают те или иные изменения, можно найти границы фонем или слогов с надежностью порядка 75—80%.

Так, Реди (Reddy, 1966, 1967) показал, что сегментация текущей речи на основе выделения моментов изменения интенсивности и частоты пересечения нулевой линии суммарного сигнала позволяет построить общую программу распознавания текущей речи на вычислительной машине. Надежность членения в такой системе составила 81% фонем из общего числа предъявленных 287.

В работе Тафтса и др. (Tufts et al., 1965) рассматривается возможность сегментации речевого потока с помощью разделения спектра речи на несколько частотных участков и слежения за величиной, получающегося после попарного вычитания текущей интенсивности огибающих напряжений на выходах соответствующих фильтров. Решение о появлении границы фонемы принималось при переходе разностного значения указанных огибающих через заданное пороговое значение. В работе использовались четыре частотных участка и двуполярные пороговые решающие схемы. На двух тестовых фразах по данным 4 дикторов авторы получили надежность распознавания границ фонем около 75%.

В ряде других исследований для сегментации использовались данные о скорости изменения энергии в тех или иных участках спектра речевого сигнала (Liljencrants, 1963; Sakai, Doshita, 1963; Focht, 1964), применялся способ анализа через синтез (Inomata, 1962), а также испытывались смешанные признаки (Stowe, 1963).

Во всех случаях авторам удавалось на ограниченном речевом материале получать сегментацию на фонемы и слоги с надежностью от 75 до 90% соответственно. Психоакустические и лингвистические исследования (Дукельский, 1962; Ohmen, 1962) также указывают на принципиальную возможность сегментации речевого потока на отрезки, соответствующие элементарным звукам речи.

Упомянутые выше способы сегментации речевого потока основывались на выделении формальных физических признаков границ сегментов, без какого-либо учета процессов переработки сигналов в слуховой системе человека.

Исключением является сообщение о работе Манди и Фале (Mundie, Fahle, 1965), в которой сигнал пропусклся через аналоговую модель улитки и нервной сети с последующим выделением из получившегося распределения импульсов комбинаций, соот-

ветствующих сонорным звукам. Последние служили указателями границ сегментов текущей речи, из которых формировался алфавит для последующего использования в компиляционном синтезаторе. Однако ни вид получившихся сегментов, ни число их, ни какие-либо оценки качества синтезируемой речи в упомянутом сообщении не приводятся.

В модели восприятия речевых сигналов, предложенной в работе Бондарко и др. (1968), предполагается существование в слуховой системе специального механизма сегментации речевого потока на элементы типа открытых слогов.

Представляется интересным рассмотреть возможную структуру такого механизма подробнее, с учетом современных сведений о характере переработки информации о звуковых стимулах на периферии слуховой системы. Рядом исследователей (Гершуни, 1967; Радионова, 1967) описаны нейроны с фазическим типом ответа на стимул, фактически выделяющие лишь начало действия стимула. Тем самым основная функциональная роль, выполняемая фазическими нейронами, заключается в выделении момента времени появления стимула. Так как стимулом можно считать любое изменение интенсивности сигнала, то фазические ответы будут фиксировать моменты времени, когда эти изменения имеют место.

По данным, полученным в лаборатории физиологии слуха (Гершуни и др., 1967), нейроны с фазическим ответом наблюдаются почти на всех уровнях нервной части органа слуха. Можно полагать, что процесс выделения моментов изменения интенсивности стимула происходит независимо во всех участках спектра, выделяемых системой частотной селекции органа слуха. В каждом из таких спектральных участков действие стимула вызывает появление некоторого возбуждающего воздействия, изменяющегося во времени. Под возбуждающим воздействием на рассматриваемый нейрон, в зависимости от его расположения в той или иной части слуховой системы (например, в кохлеарном ядре или слуховом нерве), можно понимать среднюю частоту импульсаций за определенный отрезок времени, различные виды потенциалов, возникающих в кортиевом органе под воздействием стимула, и т. д.

В рассматриваемой ниже модели роль возбуждающего воздействия на фазический нейрон играет огибающая напряжения на выходе того или иного канала частотного анализатора, имитирующего систему частотной селекции слуховой системы.

Предположим, что ответ фазического нейрона связан с параметрами возбуждающего воздействия следующим соотношением:

$$\int_0^t u_1'(\tau) e^{-\frac{t-\tau}{T}} d\tau \geq \delta. \quad (1)$$

Здесь

$$u_1'(t) = u'(t), \text{ если } u'(t) \geq u_0';$$

$$u_1'(t) = 0, \text{ если } u'(t) < u_0';$$

δ — порог возбуждения.

Решение уравнения (1) относительно времени t дает момент возбуждения модели, отсчитанный от условного начала отсчета. Сформулированный выше алгоритм срабатывания модели фазического нейрона эквивалентен предположению о том, что рассматриваемая модель реагирует на скорость изменения возбуждающего воздействия. При этом вводится порог чувствительности по скорости. Формула (1) фактически отражает закон временной суммации возбуждающего воздействия модели нейрона.

Рассмотрим зависимость характера ответа модели фазического нейрона от интенсивности и крутизны нарастания возбуждающего воздействия.

Пусть последнее описывается формулой

$$u = u_{01} (1 - e^{-\alpha t}).$$

Здесь u_{01} — стационарная величина возбуждающего воздействия, достигаемая при $t \rightarrow \infty$; α — коэффициент, определяющий скорость нарастания возбуждающего воздействия.

Из формулы (1) видно, что модель срабатывает, если возбуждающее воздействие меняется достаточно быстро: $u'(t) \geq u_0'$ и участок быстрого изменения имеет определенную протяженность. Приближенно можно считать, что модель фазического нейрона срабатывает в том случае, если возбуждающее воздействие достигнет некоторого порогового значения δ не позже, чем за заданный промежуток времени T_0 .

В каждой частотной полосе, в которой появление стимула вызывает быстрое нарастание возбуждающего воздействия, создаются условия для возникновения ответов фазических нейронов.

Вместе с тем последние не будут возбуждаться от коротких случайных флюктуаций возбуждающего воздействия и медленных его изменений.

На рис. 1 схематически показана зависимость моментов времени срабатывания модели фазического нейрона от интенсивности возбуждающего воздействия и скорости его нарастания.

По оси ординат отложено значение интеграла s , стоящего в левой части уравнения (1). Там же отмечено значение порога $\delta=0.7$, при котором происходит срабатывание модели.

При построении диаграммы было положено $T \gg \frac{1}{\alpha}$. Кружками на кривых отмечены моменты времени, при которых скорость изменения возбуждающего воздействия становится меньше некоторой пороговой величины ($u' \leq u_0' = 0.3$). Начиная с этого момента $s = \text{const}$ (для заданного значения u_0'), кривые, продол-

женные на рисунке прерывистой линией, соответствуют случаю, когда $u'_0=0$.

При принятых значениях параметров и $\delta=0.7$ возбуждающие воздействия, описанные кривыми 1 и 2, не приведут к срабатыванию модели фазического нейрона.

Последний отвечает лишь на возбуждающие воздействия вида 3, 4, 5 (моменты времени t_3, t_4, t_5).

При условии $u'_0=0$ срабатывание будет наблюдаться, кроме моментов времени t_3, t_4, t_5 , также и при $t=t_{n_1}$ и $t=t_{n_2}$.

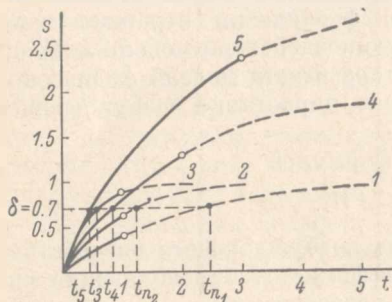


Рис. 1. Диаграмма, иллюстрирующая срабатывание модели фазического нейрона при возбуждающем воздействии вида $u=u_{10}(1-e^{-\alpha t})$.

Объяснение в тексте.

Легко видеть, что зависимость величины задержки ответа модели фазического нейрона ($u'_0 \neq 0$) от параметров α и u_{01} значительно меньше таковой у схемы, срабатывающей при достижении возбуждающим воздействием пороговой величины независимо от скорости его нарастания (т. е. при $u'_0=0$).

Можно предположить, что решение о наличии или отсутствии границы сегмента формируется в общем решающем устройстве, воспринимающем информацию о фазических ответах во всех частотных каналах анализирующей части слуховой системы. При этом необходимо иметь в виду, что при формировании фазических ответов в канале следует учитывать такие общие для нервных сетей особенности, как торможение слабого сигнала сильным, действующее в пределах рассматриваемого частотного канала и в течение определенного отрезка времени.

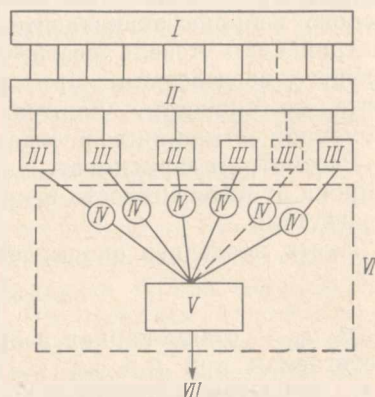


Рис. 2. Функциональная схема модели гипотетического механизма сегментации в слуховой системе.

I — модель основной мембраны и рецепторной системы; II — механизм обострения частотной избирательности основной мембраны; III — выделители моментов резкого изменения энергии сигнала в данной частотной полосе; IV — весовые коэффициенты; V — решающая схема; VI — адаптивная схема совпадений; VII — импульсы сегментации, используемые при опознавании сигналов.

На основании сказанного можно представить функциональную схему механизма сегментации потока сигналов так, как изображено на рис. 2.

Существенным свойством функциональной схемы, приведенной на рис. 2, является возможность ее подстройки под конкретный поток сигналов, подлежащий сегментации.

В принципе перестраиваемыми параметрами могут являться: пороги δ и u'_0 в отдельных частотных каналах *III*, весовые коэффициенты *IV*, правило работы решающей схемы *V*.

Можно предполагать, что в процессе распознавания нового ансамбля сигналов происходит подстройка указанных параметров механизма сегментации таким образом, чтобы выделяемые отрезки опознавались наилучшим образом (с наибольшей надежностью).

На рис. 2 решающая схема и весовые коэффициенты объединены в адаптивную схему совпадений. В таком варианте адаптивность схемы заключается в возможности изменения весов и правила совпадения до тех пор, пока последовательность окончных импульсов на выходе механизма сегментации не будет совпадать с заданной извне (например, из высших отделов мозга или со стороны «учителя»).

Последовательность операций, выполняемая для выработки параметров механизма сегментации, используемых для членения заданного потока сигналов, определяется алгоритмом обучения механизма.

В свою очередь последовательность операций, выполняемых над потоком сигналов при его сегментации после конца обучения, будем называть алгоритмом работы механизма сегментации. В последний входит также правило выработки первичных импульсов сегментации, т. е. задание величин δ и u'_0 для каждого частотного канала *III*.

Имеет смысл рассматривать целый ряд алгоритмов выделения импульсов первичной сегментации и выработки на их основе итоговых решений (импульсов) о границах сегментов. Окончательное сравнение таких алгоритмов представится возможным вести лишь в будущем, на основе обработки большого количества материала и учета данных еще не поставленных прямых электрофизиологических экспериментов. В данной работе приведен пример использования простейшего варианта алгоритма работы механизма сегментации.

Предположим, что в каждом частотном канале анализирующей части слуховой системы имеется ансамбль физических нейронов. При формировании физических ответов необходимо задаться двумя параметрами: порогом срабатывания δ и временем суммации T_0 . Величина T_0 определяется в свою очередь значением порога чувствительности модели u'_0 и постоянной времени T .

Примем максимальное время суммации равным 30 мсек. (исходя из характеристик ответов реальных физических нейронов).

Возбуждающее воздействие на выходе каналов анализирующей части слуховой системы считаем инвариантным к изменению интенсивности. Такое предположение является оправданным, в частности, если считать, что формирование возбуждающего воздействия осуществляется пропорционально разности логарифмов от величины отклонений смежных точек основной мембраны от положения равновесия. Пусть в результате обучения в каждом частотном канале устанавливается свой порог срабатывания. В частном случае величина такого порога может оцениваться

№ канала	$f_{\text{рез}}$	$f_{\text{н}}$	$f_{\text{п}}$	Δf	K_{γ}
1	202	179.5	227	47.5	2
2	403	359	453	44	4
3	640	571	718	147	7
4	806	718	906	188	7
5	1280	1141	1436	295	20
6	1614	1436	1812	376	40
7	2030	1812	2220	408	10
8	2560	2280	1870	490	20
9	3220	2870	3620	750	40

по отношению к максимальному выбросу возбуждающего воздействия на выходе рассматриваемого канала. Предположим, что

для i -го канала $\delta_i = \frac{1}{2} u_{i \text{ max}}$,

где $u_{i \text{ max}}$ — максимальный выброс возбуждающего воздействия у i -го канала. Тогда правило для определения моментов срабатывания модели физического нейрона будет следующим: модель срабатывает в том случае, если за время, не боль-

шее чем 30 мсек. от точки отсчета, возбуждающее воздействие достигает значения $\delta_i = \frac{1}{2} u_{i \text{ max}}$. Если такое событие произошло, то в качестве следующей начальной точки отсчета выбирается момент времени, когда произошло срабатывание модели. Если срабатывания не произошло, то начальная точка временного отсчета смещается на 5 мсек. по отношению к предыдущей.

Импульсы физических ответов, выработанные в отдельных каналах, поступают на общую решающую схему. Последняя выдает импульсы конечной сегментации в моменты времени, когда по крайней мере с двух каналов одновременно поступают импульсы первичной сегментации.

Одновременным считается поступление импульсов в пределах интервала времени совпадения (30 мсек.). Последовательность импульсов, разделенных интервалами менее 30 мсек., считается одним импульсом. Время поступления такого импульса считается по первому из пачки.

В соответствии со сформулированным выше правилом были сформированы импульсы сегментации для фразы «Вот приду и сяду». Сама фраза записывалась на магнитофон, соединенный с входом анализатора спектра типа СЗЧ через усилитель с автоматической регулировкой выходного напряжения глубиной около 40 дБ. В анализаторе СЗЧ использовались 9 частотных каналов. Централь-

ные частоты $f_{\text{рез}}$, нижняя $f_{\text{н}}$ и верхняя $f_{\text{в}}$ границы полос пропускания, а также полосы пропускания Δf анализатора СЗЧ приведены в таблице.

Коэффициент усиления K_y в каждом канале подбирался с учетом получения достаточного выходного напряжения для неискаженной записи. Для получения огибающих на выходах каналов использовался детектор с постоянной времени нагрузки около 20 мсек. Запись огибающей велась с помощью шлейфного осциллографа типа 950-1. Использовались шлейфы с диапазоном частот 0—4500 гц.

На рис. 3 приведены огибающие напряжения на выходах каналов анализатора (1—9) и отмечены моменты срабатывания модели фазического нейрона. Первая кривая сверху (0) является огибающей суммарного сигнала. Были использованы три значения порогов:

$$\delta_{i_1} = \frac{u_{i_{\text{max}}}}{2}; \delta_{i_2} = \frac{2}{5} u_{i_{\text{max}}}; \delta_{i_3} = u_{i_{\text{min}}} \approx \frac{1}{5} u_{i_{\text{max}}}.$$

В строках 1а, 2а, ..., 9а приведены моменты первичной сегментации, выделенные при условии $\delta_i \approx \frac{1}{5} u_{i_{\text{max}}}$, где i — номер канала. В строках 1б, 2б, ..., 9б даны моменты первичной сегментации, выделенные при условии $\delta_i = \frac{2}{5} u_{i_{\text{max}}}$. И, наконец, линии 1в, 2в, ..., 9в представляют моменты первичной сегментации, выделенные при условии $\delta_i = \frac{1}{2} u_{i_{\text{max}}}$.

Влияние правила работы решающей схемы системы совпадения на структуру потока импульсов сегментации видно из сравнения строк 11, 12, 13 на рис. 3. На этих строках приведены: моменты времени появления импульсов сегментации, построенные на основе правила одновременного существования импульсов первичной сегментации не менее чем в пяти каналах ($\delta_i \approx \frac{1}{5} u_{i_{\text{max}}}$) (строка 11); моменты времени появления импульсов сегментации, построенные на основе правила одновременного существования импульсов первичной сегментации не менее чем в двух каналах (для $\delta_i = \frac{2}{5} u_{i_{\text{max}}}$) (строка 12); моменты времени появления импульсов сегментации, построенные на основе правила одновременного существования импульсов первичной сегментации хотя бы в одном канале (для $\delta_i = \frac{1}{2} u_{i_{\text{max}}}$) (строка 13).

Для сравнения структуры сегментационных импульсов на строке 14 даны границы сегментов, определенные по видеограмме.

Масштаб времени на рис. 3 представлен в строке 10, где интервал между двумя соседними импульсами равен отрезку времени 0.02 сек. Последняя строка рис. 3 показывает истинную сегментацию исследуемой фразы на фоне. Сравнение четырех нижних последовательностей моментов сегментации показывает,

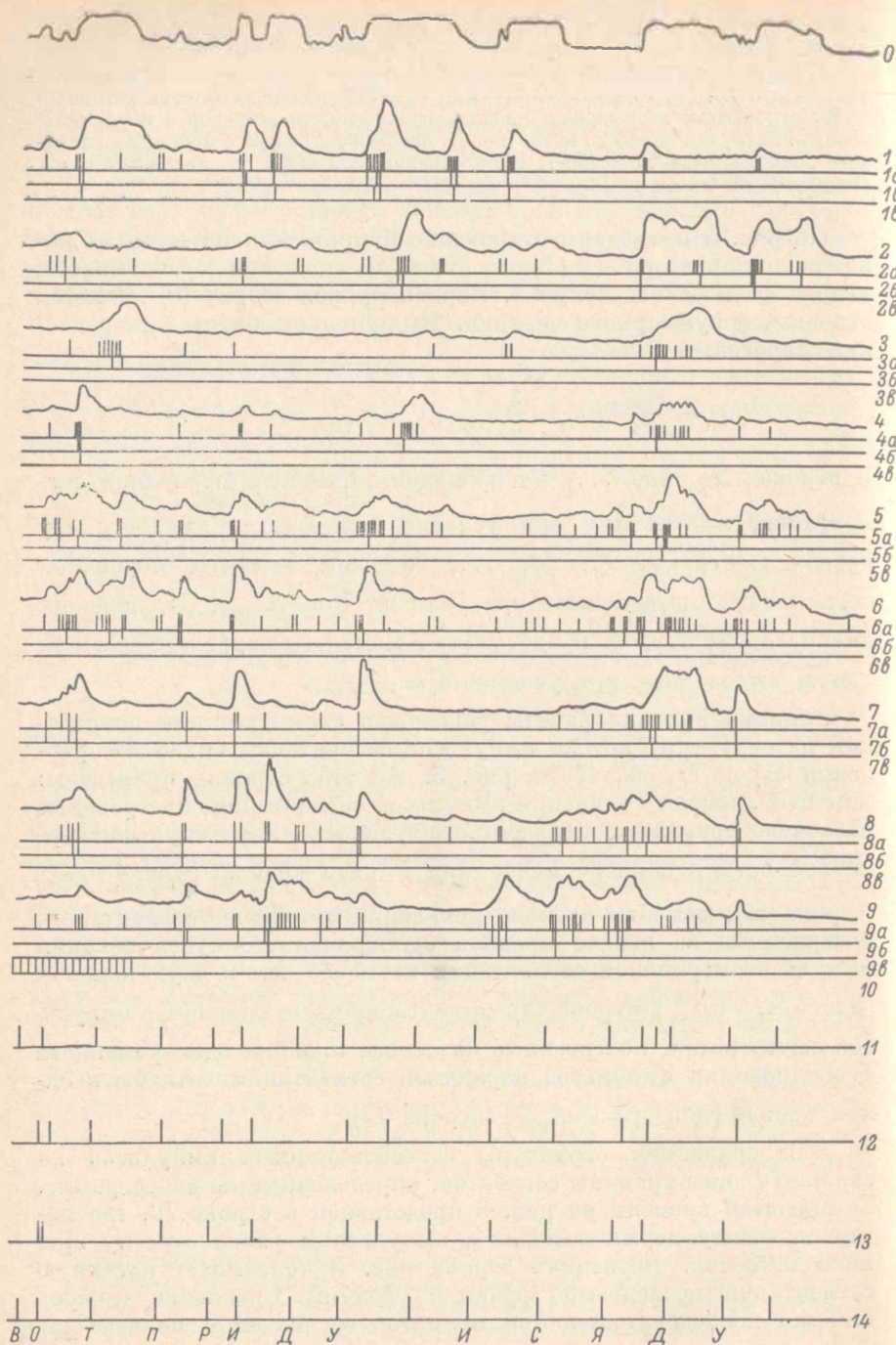


Рис. 3. Диаграмма сегментации фразы *Вот придут и сядут*.
Объяснение в тексте.

что использование ответов фазических нейронов в качестве первичных данных для сегментирования текущей речи на естественные сегменты может быть весьма перспективным. Из предъявленных 13 фонем границы у 12 из них оказались отмеченными правильно.

Правило работы схемы совпадения в данном примере выбиралось исходя из разделения потока сигналов на сегменты, соответствующие фонемам. Как уже говорилось выше, вид сегментов должен определяться типом алгоритма распознавания сигналов. Если процедура распознавания включает в себя учет ритмических признаков, то следует ожидать настройку механизма сегментации на выделение отрезков, инвариантных к темпу потока сигналов (например, для речи такие сегменты, по-видимому, соответствуют открытым слогам).

Результаты проведенного эксперимента позволяют ожидать, что на основе описанного правила можно будет создать алгоритм устойчивой сегментации текущей речи на естественные элементы, свойства которых определяются их объективными характеристиками.

В заключение автор приносит благодарность инженеру В. Н. Лапченко и лаборанту П. И. Бурухину за выполнение экспериментов в настоящей работе.

Л и т е р а т у р а

- Бондарко Л. В., Н. Г. Загоруйко, В. А. Кожевников, А. П. Молчанов, Л. А. Чистович. 1968. Модель восприятия речи человеком. Новосибирск.
- Гершуни Г. В. 1967. О механизмах слуха (в связи с исследованием временных и временно-частотных характеристик слуховой системы). В сб.: Механизмы слуха, Л.: 3—32.
- Дукельский Н. П. 1962. Принципы сегментации речевого потока. М.—Л.
- Радионова Е. А. 1967. О значении временных характеристик реакции нейронов кохлеарного ядра на звуковые сигналы. В сб.: Механизмы слуха, Л.: 32—49.
- Focht L. R. 1964. Speech energy rate of change as a function of its spectral location. J. Acoust. Soc. Amer., 36: 1030.
- Inomata S. 1962. Program for active segmentation and reduction of phonetic parameters. Proc. of Speech Communication Seminar, Stockholm: B7.
- Liljencrants J. 1963. A few experiments on voice-voiceless identification and time-segmentation of speech. J. Acoust. Soc. Amer., 37: 1113.
- Mundie J. R., W. A. Fahle. 1965. Bionics approach to speech segmentation and synthesis. J. Acoust. Soc. Amer., 38: 934.
- Ohmen S. 1962. Perceptual segments and rate of changes on spectrum in connected speech. Proc. of Speech Communication Seminar, Stockholm.
- Reddy D. R. 1966. Segmentation of speech sounds. J. Acoust. Soc. Amer., 40: 307.
- Reddy D. R. 1967. Computer segmentation of connected speech. J. Acoust. Soc., Amer. 42: 329.
- Sakai T., S. Doshita. 1963. The automatic speech recognition system. IEEE Trans. Electr. Commun., 12: 835.
- Stowe A. N. 1963. Segmentation of speech into syllables. J. Acoust. Soc. Amer., 35: 806.

ВОСПРИЯТИЕ УДАРЕНИЯ В ПРОСТОЙ ЗВУКОВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Н. А. Федорова

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

В соответствии с утверждениями классической психоакустики о том, что слуховое ощущение соответствует описанию сигнала набором значений таких его элементарных субъективных признаков, как высота, громкость и субъективная длительность (Чистович, 1966), большинство авторов рассматривает в качестве акустических коррелятов ударения частоту основного тона гласного, его длительность и интенсивность. Однако многочисленными экспериментами установлено, что восприятие ударенности не может быть достаточно просто и однозначно описано ни каким-то одним из этих параметров, ни их сочетанием. Акустические измерения на изолированно произнесенных словах и фразах показывают, что в ударном гласном возможны чрезвычайно разнообразные сочетания значений частоты основного тона, длительности и интенсивности.

Поэтому понятны попытки целого ряда авторов найти некоторую величину, которая, являясь функцией указанных параметров, сама могла бы быть использована в качестве параметра для описания восприятия ударения и его автоматического распознавания. В частности, в последнее время достаточно широкое распространение получило представление об энергии звука как мере ударения (Fant, 1958, 1960; Lehistе, Peterson, 1959; Lieberman, 1960).

Фант (Fant, 1958, 1960) впервые предложил рассматривать в качестве слухового коррелята ударения величину площади, ограниченной кривой, характеризующей изменение интенсивности на протяжении слога. При квадратичном характере выпрямления эта величина имеет размерность энергии. Целесообразность использования такого параметра обосновывается Фантом «с одной стороны, данными о том, что громкость зависит не только от интенсивности, но и от длительности звука, с другой же стороны, опытами по анализу и синтезу речи, согласно которым одного только укорочения длительности достаточно для изменения восприятия ударенности» (Фант, 1964 : 222).

Либерман (Lieberman, 1960) использовал величину интеграла амплитуды по времени в алгоритме автоматического выделения ударных слогов. В использованном им экспериментальном материале (двусложные квазиомонимы) интеграл ударного слога (подсчитанный графически) в 92% случаев был больше интеграла неударного слога того же слова.

Использование в качестве коррелята ударения именно отношения энергий является вполне оправданным. Постоянство диф-

ференциальных порогов по интенсивности (в дБ) по всей шкале амплитуд говорит о том, что человек при слуховом восприятии использует именно отношения интенсивностей, а не их разность. Опыты по восприятию удара в синтезированных двусложных словах (Fry, 1955) также подтверждают, что для человека важны именно отношения длительностей и интенсивностей гласных, а не их разности. А так как звуковая энергия определяется обеими этими величинами, то естественно предположить, что и в этом случае важны будут именно отношения звуковых энергий.

Поскольку известна большая значимость энергии как акустического параметра для выделения ударного слога (Lieberman, 1960), представляло интерес проверить, насколько существенным окажется этот параметр при решении аналогичной задачи человеком.

Насколько нам известно, попытки экспериментально проверить роль энергии при идентификации удара в опытах по восприятию не предпринимались.

Предположение о том, что ударение определяется отношением звуковых энергий, влечет за собой следующие допущения: 1) человек способен измерять энергию звука; 2) человек может осуществлять сравнение соседних элементов.

С точки зрения психоакустики, эти допущения не противоречат имеющимся данным. Известно, что слуховая система осуществляет интегрирование сигнала по времени, что подтверждается данными о понижении порогов восприятия при увеличении длительности стимула (Wright, 1964a, 1964b). На основании этих данных, а также данных о зависимости громкости сигнала от его длительности Цвишлоцким (Zwislocki, 1960) была предложена модель, описывающая явление временной слуховой суммации.

Возможность измерения энергии подтверждается также электрофизиологическими данными. Известно существование нейронов с тоническим разрядом, ответы которых характеризуются значительным понижением порогов при увеличении длительности сигнала, т. е. выраженным явлением временной суммации (Гершуни, 1967).

Что касается второго допущения, то известно, что в условиях опыта по парному сравнению сигналов, разделенных небольшим промежутком времени, человек ведет себя так, как будто располагает чрезвычайно подробным описанием сигналов. В опытах с абсолютным распознаванием он ведет себя так, как если бы сигнал описывался чрезвычайно грубо и неточно (Чистович, 1966).

Предположение об использовании человеком звуковой энергии как слухового параметра удара решено было проверить на примере восприятия удара в простой звуковой последовательности, состоящей из двух одинаковых гласных, разделенных небольшим временным интервалом.

Таким образом, гипотеза, подлежащая проверке, формулируется следующим образом: суждение о месте ударения в двусложном слове (при постоянной частоте основного тона) выносится человеком на основании измерения отношения величин энергий первого и второго слога. Если это отношение больше единицы, ударным считается первый слог, если меньше — второй.

МЕТОДИКА

Для проверки гипотезы были проведены эксперименты по определению места ударения в синтезированных звуко сочетаниях, которые воспринимаются на слух как два гласных [а], разделенных согласным звуком.

Звуко сочетания синтезировались на специальной установке, блок-схема которой приведена на рис. 1. В качестве источника сигнала использовался генератор сложных напряжений ГСН-1.

Значения параметров на генераторе подбирались таким образом, чтобы полученный сигнал был близок по звучанию к русскому гласному [а]. Частота основного тона сигнала была равна 200 гц. Интенсивность второй и последую-

щих гармоник относительно интенсивности основного тона составляла соответственно: 0,0, 0,0, 1,0, 0,5, 1,0.

С помощью двухканального электронного ключа формировались две посылки сигнала, длительность каждой из которых, как и длительность интервала между ними, задавалась блоком временных задержек. Интенсивность каждой посылки регулировалась отдельным аттенуатором (со шкалой в неперах). После суммирования сегментов сигналы подавались на магнитофон МЭЗ-28. Время

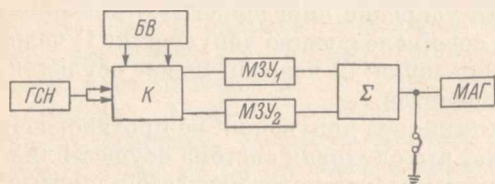


Рис. 1. Блок-схема установки для синтеза двусложного звуко сочетания [ага].

ГСН — генератор сложных напряжений; К — двухканальный электронный ключ; БВ — блок временных задержек; МЗУ — аттенуаторы; Σ — сумматор; МАГ — магнитофон.

нарастания интенсивности каждого из сегментов было в пределах 10 мсек.

Для эксперимента был синтезирован набор стимулов, составленных из двух отрезков гласного [а], следующих друг за другом с интервалом 30 мсек. Такого рода стимулы воспринимаются на слух как двусложные сочетания [ага].

Длительность каждого сегмента стимула менялась от 200 до 100 мсек. ступенями по 20 мсек. (всего было 6 градаций).

Интенсивность каждого из сегментов менялась ступенями по 2,6 дб в диапазоне 0—23,4 дб (10 градаций) относительно интенсивности исходного стимула.

Значение уровня интенсивности обоих сегментов исходного стимула было одинаковым (максимальным) и подбиралось так, чтобы оно попадало в зону слухового комфорта, а самый слабый сигнал (—23,4 дб относительно максимального уровня) был бы слышен при этом достаточно ясно. Длительность первого и второго сегментов исходного стимула была равна 200 мсек.

Все прочие стимулы были получены из исходного путем изменения длительности и интенсивности каждого из сегментов при соблюдении следующих правил:

- 1) одновременно в каждом сегменте стимула может быть изменен, по сравнению с исходным, только один параметр;
- 2) если в первом сегменте меняется длительность, то во втором сегменте меняется интенсивность, и наоборот;

3) и первый и второй гласный звуко сочетания порознь подвергаются одинаковым изменениям.

Таким образом было получено 120 разных стимулов. В 59 из них интенсивность первого сегмента I_1 и длительность второго t_2 имели постоянные и максимальные значения, т. е.

$$I_1 = I_{\max} = \text{const} > I_2;$$

$$t_1 < t_2 = t_{\max} = \text{const}.$$

В 59 других — постоянными и максимальными были значения интенсивности второго сегмента I_2 и длительности первого t_1 , т. е.

$$I_1 < I_2 = I_{\max} = \text{const};$$

$$t_1 = t_{\max} = \text{const} > t_2.$$

Кроме того, в набор входили два исходных стимула, у которых

$$I_1 = I_2 = I_{\max};$$

$$t_1 = t_2 = t_{\max}.$$

Отсюда видно, что стимулы, в которых один из сегментов превышал бы другой одновременно и по длительности и по интенсивности, не исследовались.

Стимулы были записаны на магнитную ленту в случайном порядке. Интервал между ними составлял приблизительно 4 сек.

После записи на магнитофон было проверено соответствие полученных значений параметров стимулов задаваемым. Длительность сегментов стимулов измерялась по огибающей уровня интенсивности, записанной с магнитной ленты на чернилопишущем осциллографе (точность измерений составляет 4 мсек.). Уровень интенсивности измерялся по записи, полученной на самописце уровня напряжения Н-110 (с точностью 0.5 дБ). Проверка показала, что реально полученные значения разности уровней интенсивности первого и второго гласных соответствуют заданным. Значения длительностей несколько отличались от заданных, что было учтено при последующей обработке полученного материала.

Испытуемые прослушивали стимулы в звукозаглушенной камере через головные телефоны ТД-6. Задача испытуемых состояла в том, чтобы определить место удара и записать в протоколе цифру 1, если ударение падает на первое [а], и цифру 2, если на второе. В работе использовался метод форсированного выбора: ответы «не знаю» не разрешались.

В опытах принимали участие 16 испытуемых, каждый из которых прослушал таблицу по несколько раз. Всего на каждый из 120 стимулов было получено по 50 ответов. При обработке материала для каждого стимула подсчитывался процент восприятия удара на каждом из слогов (гласных).

РЕЗУЛЬТАТЫ

Как уже было сказано, мы предполагаем, что в качестве меры ударенности используется отношение энергии первого слога к энергии второго. Так как форма огибающей интенсивности была близка к прямоугольной, отношения энергий приближенно могут быть записаны в виде

$$\frac{I_1 t_1}{I_2 t_2} = \frac{P_1^2 t_1}{P_2^2 t_2}, \quad (1)$$

где I_1 и I_2 — интенсивность первого и второго гласного; t_1 и t_2 — длительность первого и второго гласного; P_1 и P_2 — значения звукового давления.

Тогда вероятность восприятия удара, например, на первом слоге, должна быть монотонно-возрастающей функцией отношения энергий (1) или монотонно-возрастающей функцией величины $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$, где первый член представляет собой разность между уровнями интенсивности первого и второго гласного (в дБ).

На рис. 2 и 3 показана зависимость частоты восприятия удара на первом слоге h_1 от величины $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$, отражающей отношение энергии первого слога к энергии второго слога. Параметром кривых, изображенных на рис. 2, является величина $20 \lg \frac{P_1}{P_2}$; параметр кривых на рис. 3 — величина $10 \lg \frac{t_1}{t_2}$.

Как видно из распределения ответов, полученная зависимость h_1 от величины $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$ не может быть описана монотонно-возрастающей функцией. Точки, имеющие одинаковые значения h_1 , расположены вдоль всей оси абсцисс; для стимулов, которые имеют равные или очень близкие значения отношений энергий, величина h_1 также неодинакова и меняется в очень большом диапазоне (например, в одном из случаев от 2 до 93%).

Если бы восприятие удара определялось отношением звуковых энергий, мы имели бы не семейство кривых, а одну плавно возрастающую кривую. Как видно из рисунков, этого не наблюдается. Кривые имеют значительный разрыв. Все это свидетельствует о несоответствии полученных данных выдвинутой гипотезе, согласно которой решение об ударе принимается человеком на основании сравнения величин энергий первого и второго слогов.

Левая часть обоих рисунков достаточно заметно отличается от правой. В левой части, где выполняются условия $t_1 > t_2$ и $I_1 < I_2$, распределение точек несколько больше соответствует предложенной гипотезе. Частота восприятия удара на первом слоге с возрастанием величины $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$ увеличивается достаточно равномерной полосой (рис. 3). Однако ширина этой полосы — диапазон значений h_1 в ответ на стимулы с одинаковым отношением энергии первого слога к энергии второго слога — достаточно велика. Это также противоречит проверяемой гипотезе.

Для того чтобы проверить, согласуются ли наши результаты с результатами, полученными другими исследователями, дополнительно была проведена обработка данных, полученных Ясесом, Мортон и Штеффен-Батог на аудиторах-англиканах (Morton, Jassem, 1965) и аудиторах-поляках (Jassem et al., 1968). В экспериментах этих авторов слушатели должны были определить место удара в синтезированных словах [sasa], [sɔsɔ], [sisi], в которых определенным образом менялись значения час-

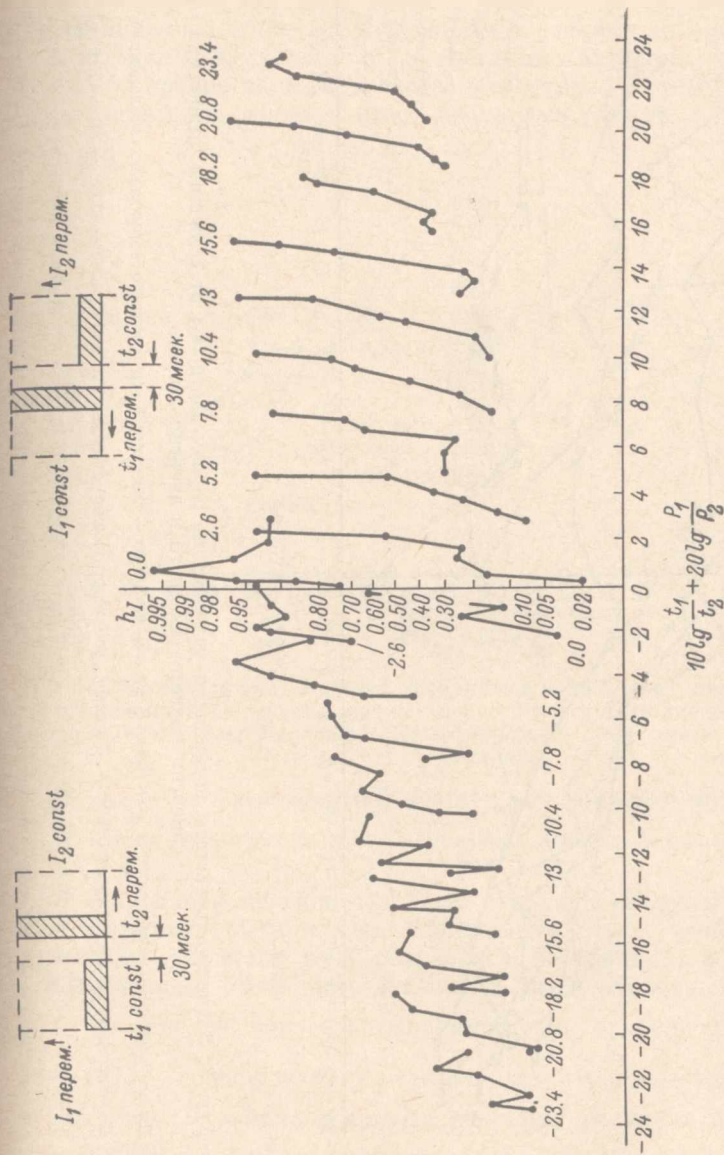


Рис. 2. Зависимость частоты восприятия удара на первом слоге (h_1) от величины отношения энергии первого гласного к энергии второго.

По оси абсцисс — величина $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$; по оси ординат — значения h_1 (в масштабе нормальных отклонений).

Параметры кривых — величина $20 \lg \frac{P_1}{P_2}$.

Вверху показаны схемы изменения длительности и интенсивности экспериментальных стимулов.

тоты основного тона, длительности и интенсивности. Так как в этих экспериментах слушателям было предложено три рода ответов (ударение на первом слоге, на втором слоге и ответ «не знаю», когда они затруднялись сделать выбор между слогами), при проверке гипотезы по данным указанных авторов за h_{I_i} была взята сумма двух величин, приведенных авторами: частоты суждений «ударение на первом слоге» в ответ на i -й стимул и $1/2$ частоты суждений «не знаю» в ответ на тот же стимул.

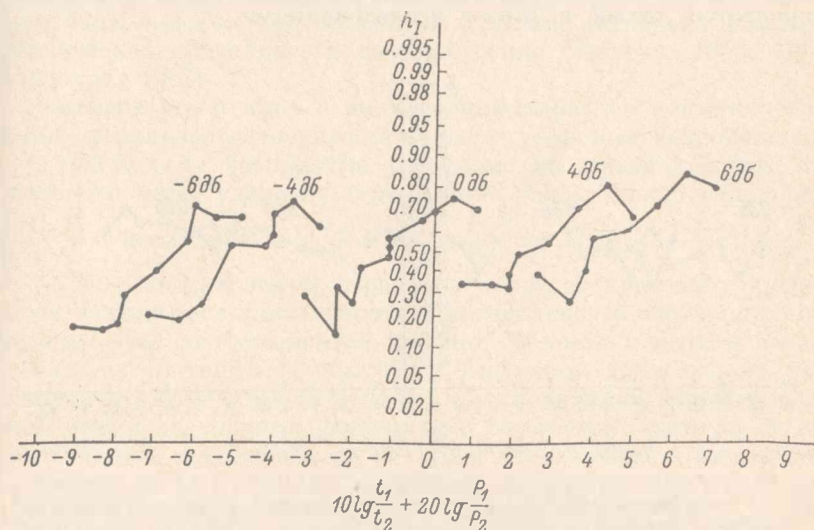


Рис. 4. Зависимость частоты восприятия ударения на первом слоге (h_I) от величины отношения энергии первого гласного к энергии второго (по данным Яссема, Мортон и Штеффен-Батог, полученным на аудиторных поляках).

По оси абсцисс — величина $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$; по оси ординат — значения h_I (в масштабе нормальных отклонений). Параметр кривых — величина $20 \lg \frac{P_1}{P_2}$.

Для проверки гипотезы были отобраны стимулы, аналогичные использованным в нашем эксперименте, т. е. такие, в которых частота основного тона оставалась постоянной, а переменными параметрами были длительность и интенсивность.

На рис. 4 показана зависимость h_I от величины $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$ для аудиторов-поляков, а на рис. 5 — для аудиторов-англичан. Параметром кривых на обоих рисунках является величина $20 \lg \frac{P_1}{P_2}$.

Так же, как и в нашем эксперименте, точки с одинаковой частотой восприятия ударения на первом слоге расположены вдоль

всей оси абсцисс. При этом в случае аудиторов-англичан все точки с небольшим разбросом лежат около 50%-й прямой (рис. 5). Это означает, что испытуемые при восприятии удара в данных стимулах ведут себя случайным образом и не отдают предпочтения ни одному из слогов, независимо от величины отношения энергий гласных в этих стимулах.

Таким образом, гипотеза об использовании человеком отношения звуковых энергий для идентификации удара не подтверждается также и этими исследованиями.

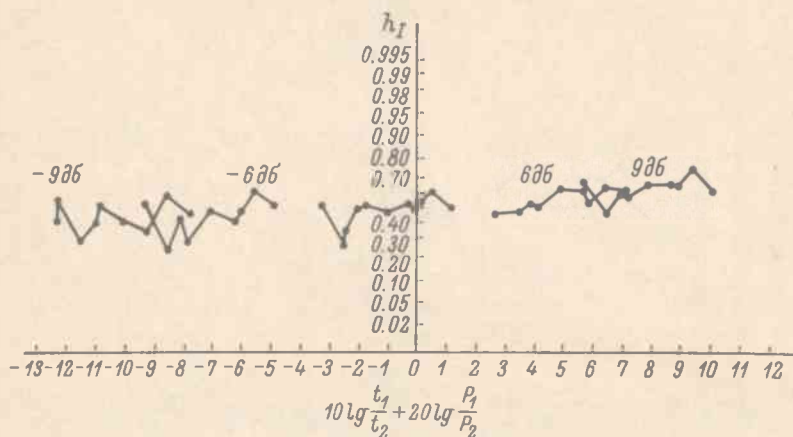


Рис. 5. Зависимость частоты восприятия удара на первом слоге (h_I) от величины отношения энергии первого гласного к энергии второго (по данным Яссема и Мортон, полученным на аудиторах-англичаных).

По оси абсцисс — величина $20 \lg \frac{P_1}{P_2} + 10 \lg \frac{t_1}{t_2}$; по оси ординат — значения h_I (в масштабе нормальных отклонений). Параметр кривых — величина $20 \lg \frac{P_1}{P_2}$.

Качественная оценка данных, полученных в нашем эксперименте (рис. 2 и 3) позволяет считать, что восприятие удара в значительно большей степени определяется отношением длительностей гласных, чем разностью их уровней интенсивности. Как видно из рисунков, в выбранном масштабе и выбранных координатах кривые, показывающие зависимость восприятия удара на первом слоге от логарифма отношения длительностей гласных (параметр этих кривых $20 \lg \frac{P_1}{P_2}$, рис. 2), имеют более крутой наклон, чем кривые, показывающие зависимость восприятия удара от разности между уровнями интенсивности первого и второго гласных (параметр этих кривых $10 \lg \frac{t_1}{t_2}$, рис. 3).

Кроме того, при сравнении правой и левой частей рис. 3 видно, что зависимость от разности уровней интенсивности сильнее

выражена в левой части рисунка (наклон этих кривых является большим). При этом легко заметить, что наклон кривых в левой части этого рисунка носит также и более стабильный характер, чем в правой, где он достаточно заметно меняется от одной кривой к другой.

Зависимость от логарифма отношения длительностей (рис. 2), наоборот, кажется более выраженной в правой части.

Известно, что и громкость и субъективная длительность звука могут быть описаны как степенные функции звукового давления и физической длительности соответственно (Stevens, 1955, 1961; Гольдбурт, 1967).

Учитывая это и имея в виду разный характер влияния отношения длительностей и разности между уровнями интенсивности двух гласных на восприятие удара, мы предположили, что в качестве меры удара может быть использовано отношение

$$\frac{P_1^b \cdot t_1^a}{P_2^b \cdot t_2^a}, \text{ или величина } a20 \lg \frac{t_1}{t_2} + b20 \lg \frac{P_1}{P_2} \left(\text{или } 20 \lg \frac{t_1}{t_2} + \frac{b}{a} 20 \lg \frac{P_1}{P_2} \right),$$

где a и b — коэффициенты, отражающие при восприятии удара относительные веса длительности и интенсивности соответственно. Введение этих коэффициентов придает гибкость модели и делает ее более адекватной. Из опытов по имитации, абсолютному распознаванию и психофизическому шкалированию известно, что связь между значениями физического параметра стимула и его субъективным отражением не является однозначной и изменяется в зависимости от диапазона изменений параметра стимула и других факторов (Клаас, 1959; Голузина и др., 1966).

Для вычисления значений коэффициентов a и b было использовано допущение, что частота восприятия удара на первом слоге (h_1), изображенная в масштабе нормальных отклонений, является линейной функцией величины $20 \lg \frac{t_1}{t_2}$ или $20 \lg \frac{P_1}{P_2}$.

Проверить это допущение на материале данной работы не представлялось возможным в связи с очень большим разбросом точек. Возможно, что этот разброс был связан с влиянием контекста (в нашей таблице каждый стимул встречался только в одном положении). Поэтому для проверки этого допущения были проведены дополнительные эксперименты по восприятию удара в синтезированных звукосокращениях, где переменным параметром была только длительность гласных (использовались разные значения $\frac{t_1}{t_2}$). Оказалось, что линейная аппроксимация в области от $h_1 = 0.1$ до $h_1 = 0.9$ является возможной.

Для вычисления значения коэффициента a были построены семейства зависимостей h_1 от величины $20 \lg \frac{t_1}{t_2}$ для стимулов с одинаковой разностью уровней интенсивности (в диапазоне $h_1 = 0.1 \div 0.9$). Зависимости аппроксимировались прямыми от-

дельно для стимулов, где $t_1 < t_2$, и для стимулов, где $t_1 > t_2$. Затем для каждого семейства определялся средний наклон этих прямых. Оказалось, что для стимулов, где $t_1 > t_2$, коэффициент $a=0.67$; для стимулов, где $t_1 < t_2$, $a=0.94$. Аналогичным образом были вычислены значения коэффициента b : были построены семейства зависимостей h_1 от величины $20 \lg \frac{P_1}{P_2}$ для стимулов с одинаковым отношением длительностей. Для стимулов, где $I_1 < I_2$, коэффициент b оказался равным приблизительно 0.25; для стимулов, где $I_1 > I_2$, $b=0.12$.

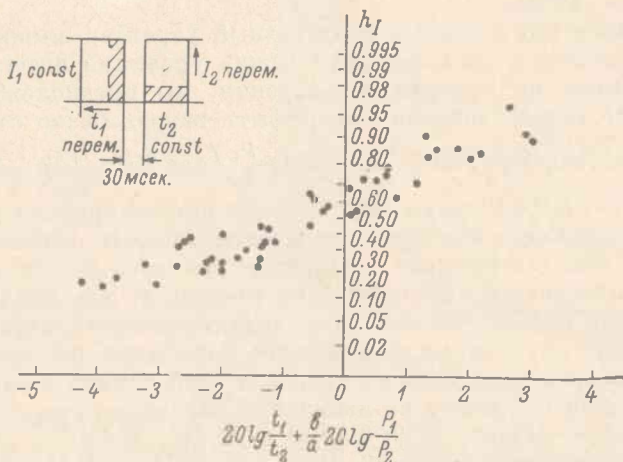


Рис. 6. Зависимость частоты восприятия ударения на первом слоге (h_1) от величины $20 \lg \frac{t_1}{t_2} + \frac{b}{a} 20 \lg \frac{P_1}{P_2}$ для стимулов, изменяющихся по схеме, изображенной в верхнем углу рисунка.

Как можно видеть из полученных значений коэффициентов a и b , веса длительности и интенсивности действительно различны; это различие оказывается большим для стимулов, у которых $I_1 > I_2$ и $t_1 < t_2$.

На рис. 6 и 7 изображена зависимость частоты суждений об ударности первого слога от величины $20 \lg \frac{t_1}{t_2} + \frac{b}{a} 20 \lg \frac{P_1}{P_2}$. Можно видеть, что для стимулов, у которых $I_1 > I_2$ и $t_1 < t_2$, линейная зависимость h_1 от выбранного аргумента выражена достаточно очевидно (рис. 6). Для стимулов, у которых $I_1 < I_2$ и $t_1 > t_2$, эта зависимость проявляется менее отчетливо (рис. 7). Разброс точек для этого случая оказался достаточно большим, однако он значительно меньше, чем в том случае, когда в качестве аргумента было взято отношение энергий гласных, или отношение длительностей, или отношение интенсивностей,

В работе не ставилась цель определить точно относительные веса рассматриваемых параметров для восприятия удара. Коэффициенты a и b были вычислены приблизительно. Существовало, однако, при этом то, что для разных стимулов приходится принимать разные значения $\frac{b}{a}$. Такой эффект можно было бы ожидать, если бы воспринимающая система меняла свои свойства во времени, т. е. величины, описывающие интенсивность и длительность сигнала в нервной системе, определялись бы не только физическими признаками данного сигнала, но и следующего за ним другого сигнала. Эффекты такого рода известны психоакустике. Например, известен эффект адаптации — уменьшение громкости

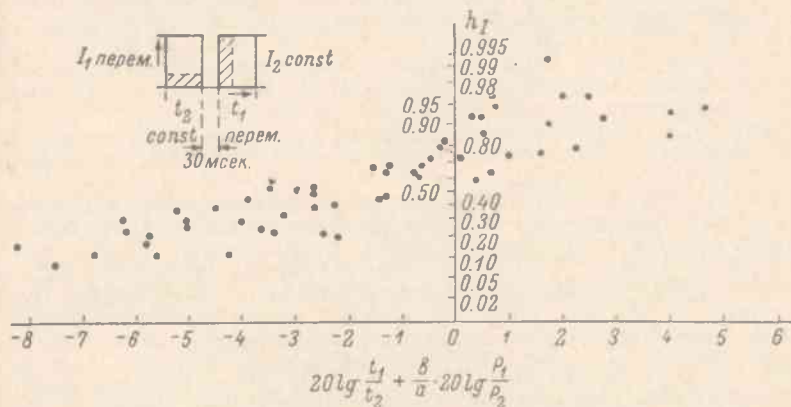


Рис. 7. Зависимость частоты восприятия удара на первом слоге (h_I) от величины $20 \lg \frac{t_1}{t_2} + \frac{b}{a} 20 \lg \frac{P_1}{P_2}$ для стимулов, изменяющихся по схеме, изображенной в верхнем углу рисунка.

сигнала одного и того же уровня при продолжающемся его действии во времени (Békésy, 1960). Гольдбурт (1967) описан эффект изменения субъективной длительности первого сигнала в последовательности, если звуки разделены коротким временным интервалом (0—30 мсек.). Длительность первого из двух звуков воспринимается отчетливо укороченной по сравнению с длительностью того же звука при одиночном его предъявлении. Вполне возможно, что разный характер зависимости h_I от величины, определяющей отношение энергий, объясняется указанными явлениями.

Таким образом, можно считать, что использование звуковой энергии в качестве слухового коррелята удара не является оправданным. Гипотеза о том, что человек принимает решение об ударе на основании величины отношения энергий первого и второго слогов (гласных), не подтверждается ни данными на-

ших экспериментов, ни данными, полученными Яссемом с соавторами на аудиторах-англичанах и аудиторах-поляках.

При суждении о месте удара человека учитывают и отношения длительностей гласных, и отношения их интенсивностей, но при этом отношения длительностей имеют значительно больший вес, чем отношения интенсивностей.

Л и т е р а т у р а

- Гершун Г. В. 1967. О механизмах слуха (в связи с исследованием временных и временно-частотных характеристик слуховой системы). В сб.: Механизмы слуха, Л.: 3—32.
- Гольдбурт С. Н. 1967. Особенности оценки человеком длительности коротких звуков. В сб.: Механизмы слуха, Л.: 150—157.
- Голузина А. Г., В. В. Люблинская, М. К. Рохтла, Л. А. Чистович. 1966. Шкалирование субъективных расстояний как метод исследования слухового восприятия. В сб.: Механизмы речеобразования и восприятия сложных звуков, М.—Л.: 86—110.
- Класс Ю. А. 1959. Об изменении точности оценки тональных сигналов в зависимости от диапазона интенсивности и частот сигналов. Пробл. физиол. акустики, IV: 45—50.
- Чистович Л. А. 1966. Психоакустика и вопросы теории восприятия речи. В кн.: Распознавание слуховых образов. Новосибирск: 68—169.
- Békésy G. 1960. Experiments in hearing. N. Y.—Toronto—London.
- Fant G. 1958. Proc. of VIII Intern. Congr. of Linguists (1957), Oslo: 282—358.
- Fant G. 1960. Acoustic theory of speech production. 's-Gravenhage. (Русск. пер.: Г. Фант. 1964. Акустическая теория речеобразования. М.).
- Fry D. B. 1955. Duration and intensity as physical correlates of linguistic stress. J. Acoust. Soc. Amer., 27 (4): 765—768.
- Jassem W., G. Morton, M. Steffen-Batog. 1968. The perception of stress in synthetic speech-like stimuli by Polish listeners. In: Speech analysis and synthesis. Ed. by W. Jassem, Warsaw, 1: 289—309.
- Lechiste J., G. E. Peterson. 1959. Vowel amplitude and phonemic stress in American English. J. Acoust. Soc. Amer., 31 (4): 428—435.
- Lieberman Ph. 1960. Some acoustic correlates of word stress in American English. J. Acoust. Soc. of Amer., 32 (4): 451—457.
- Morton J., W. Jassem. 1965. Acoustic correlates of stress. Language a. Speech, 8 (3): 159—181.
- Stevens S. S. 1955. The measurement of loudness. J. Acoust. Soc. Amer., 27: 815—829.
- Stevens S. S. 1961. The psychophysics of sensory communication. N. Y.—London.
- Wright H. N. 1964a. Temporal summation and backward masking. J. Acoust. Soc. Amer., 36 (5): 927—939.
- Wright H. N. 1964b. Backward masking for tones in narrow-band noise. J. Acoust. Soc. Amer., 36 (11): 2217—2221.
- Zwislöcki J. 1960. Theory of temporal auditory summation. J. Acoust. Soc. Amer., 32 (8): 1046—1060.

II. ПРАВИЛА ФОНЕТИЧЕСКОЙ ИНТЕРПРЕТАЦИИ И ЗАПОМИНАЕМОЕ ОПИСАНИЕ РЕЧЕВОГО ЭЛЕМЕНТА

К ВОПРОСУ О ПРОЦЕДУРЕ РАСПОЗНАВАНИЯ ФОНЕМ

Л. А. Чистович

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

В исследовании восприятия речи выделяются две достаточно самостоятельные проблемы: определение признаков, по которым сигнал описывается слухом, и определение процедуры фонемной интерпретации, т. е. перехода от слухового описания сигнала к фонемному символу.

Если первой проблеме посвящено довольно много работ, разработка второй проблемы находится еще на такой стадии, когда не определены ни задачи экспериментального исследования, ни его методы. Установлено только, что модель, осуществляющая фонемную интерпретацию, должна обладать адаптивными свойствами (Ladefoged, Broadbent, 1957).

Человек, объясняющий свое поведение в эксперименте по фонемной идентификации речеподобных стимулов, обычно говорит, что он основывается на относительном сходстве услышанного стимула с разными фонемами, выбирает ту фонему, на которую звук больше всего похож. Абсолютная величина сходства стимула с выбранной фонемой, по мнению испытуемого, также им учитывается и определяет его впечатление о том, насколько «плохим» или «хорошим» является звук. Это приводит к мысли, что в процессе фонемной интерпретации можно выделить два последовательных этапа: на первом из них осуществляется переход от слухового описания сигнала в терминах акустических признаков к описанию сигнала в терминах его сходства с возможными фонемами; на втором этапе принимается решение — выбирается одна из фонем. Если принять такую двухэтапную модель, можно сделать предположение, что функция сходства (связь между слуховым описанием и значениями сходства с фонемами) сохраняется у взрослого человека неизменной (или меняется только в процессе обучения новому языку), а все явления подстройки осуществляются

в блоке принятия решений. Тогда задача разработки модели сводится к определению функции сходства и к выяснению структуры блока принятия решений.

В настоящее время возможны самые различные гипотезы относительно способа задания функции сходства и относительно ее вида. Например, допустимо предположение, что функция сходства задана таблично, причем для очень мелких градаций (областей) в многомерном пространстве слуховых описаний. Сами значения сходства, приписанные каждой такой области, могли бы отражать относительные частоты случаев попадания разных фонем в эту область. Однако, если эта гипотеза верна, экспериментальное определение функции сходства является непосильной задачей и лишено практического смысла.

Другое столь же возможное предположение состоит в том, что функции сходства заданы отдельно для каждого выделяемого слухом параметра речевого сигнала. Общее сходство звука с данной фонемой вычисляется путем суммирования значений, полученных по каждому из параметров. Если верна эта гипотеза, экспериментальное определение функций сходства является вполне реальной, хотя и трудоемкой задачей.

Наконец, можно сделать еще более «удобное» предположение, что функции сходства с данной фонемой не только заданы отдельно для каждого параметра, но и имеют почти прямоугольную форму. Если это так, задача определения функции сходства может оказаться достаточно простой.

Целью настоящей работы была предварительная качественная проверка этой последней гипотезы.

Для исследования был избран максимально простой случай, когда выбор проводился всего из двух фонем — [b] и [m], а число полезных параметров сигнала также равнялось двум. Стимулами служили синтетические открытые слоги, воспринимаемые как [ba] или [ma]. Их переменными параметрами были: I_4 — уровень интенсивности четвертой гармоники (составляющая с частотой 1000 гц) в спектре согласного и ΔF_0 — величина перепада основной частоты на переходе от согласного к гласному. Исходя из данных анализа естественных [b] и [m] и восприятия синтетических [b] и [m] (Fant, 1960; Kasprowski, Mikiel, 1966; Чистович, 1968), следовало ожидать, что сходство с [m] будет расти (сходство с [b] падать) с увеличением I_4 и уменьшением ΔF_0 .

В первой части работы исследовалось положение фонемной границы между [b] и [m] в пространстве указанных двух параметров. Применялся метод активного поиска границы.

Во второй части работы была сделана попытка использовать метод прямого численного шкалирования субъективных расстояний для определения величин сходства звука с фонемами по каждому из параметров.

МЕТОДИКА

Для синтеза слогов использовалась установка, состоящая из двух генераторов сложных напряжений ГСН-1, двухканального электронного ключа и магазинов затухания. Для прослушивания применялись телефоны ТД-6.

Первый звук в последовательности (согласный) длительностью 50 мсек. представлял собой сумму первой и четвертой гармоники частоты 250 гц. Уровень интенсивности первой гармоники составлял 30 дБ над средним слуховым порогом. Уровень интенсивности четвертой гармоники (1000 гц) был переменным параметром. Длительность второго звука ([а]-образного гласного) составляла 250 мсек., суммарный уровень интенсивности звука и относительные амплитуды гармоник сохранялись постоянными; основная частота гласного, а следовательно, и изменение основной частоты при переходе от согласного к гласному, была переменным параметром.

В экспериментах с активным поиском границы испытуемый сам управлял синтезатором сигналов. Он изменял значения сигнала по одному из переменных параметров, стараясь найти такое значение, где [b] переходит в [m] или [m] в [b].

В первой серии опытов испытуемый изменял основную частоту гласного; значение уровня интенсивности четвертой гармоники согласного устанавливалось экспериментатором. Всего использовалось 10 значений уровня, при каждом из них испытуемый повторял поиск границы по основной частоте 20 раз. Испытуемый устанавливал F_0 на гласном, поворачивая ручку расстройкой частоты генератора. Лимб расстройки частоты был от него закрыт во избежание возможности зрительного контроля.

Во второй серии опытов значение основной частоты на гласном устанавливалось экспериментатором (использовалось 8 различных значений). Испытуемый изменял амплитуду четвертой гармоники в спектре согласного. Поиск границы повторялся по 10 раз для каждого из 8 значений основной частоты на гласном. Испытуемый манипулировал ручкой плавной регулировки амплитуды четвертой гармоники на генераторе ГСН-1. Установленное им напряжение измерялось с помощью вольтметра.

В экспериментах по шкалированию испытуемым предъявлялась последовательность, состоящая из 145 стимулов, записанных на магнитную пленку. Эта последовательность была образована из 12 различных стимулов, она была составлена так, что каждый стимул следовал по одному разу за всеми остальными стимулами и за самим собой. Значения параметров стимулов были выбраны на основании данных, полученных в экспериментах по поиску фонемной границы. Прослушивание производилось с использованием тех же телефонов ТД-6 и при том же уровне интенсивности, что и в опытах по поиску границы.

Каждый испытуемый прослушивал тестовую последовательность четыре раза. При двух прослушиваниях он определял для каждого стимула субъективное расстояние от фонемы [b], при двух других прослушиваниях — от фонемы [m].

Все эксперименты были проведены на одних и тех же трех испытуемых.

ОПРЕДЕЛЕНИЕ ФОНЕМНОЙ ГРАНИЦЫ

Прежде чем описывать полученные результаты, рассмотрим, какие качественные эффекты следовало ожидать, исходя из принятой гипотезы.

Гипотеза состоит из двух допущений. Одно из них заключается в том, что по каждому из параметров задана зависимость между значениями параметра и величинами сходства с фонемами [b] и [m]; общее сходство сигнала с данной фонемой определяется

путем суммирования значений сходства по обоим параметрам. Второе допущение заключается в том, что функция сходства имеет такой вид, как показано на рис. 1, А. Интервал значений параметра, где сходство изменяется, является сравнительно узким; вне этого интервала сходство с данной фонемой равно или нулю, или некоторому максимальному значению.

Предположим, что человек, принимая решение, сравнивает суммарное сходство сигнала с фонемой [m] с суммарным сходством его с фонемой [b]; выбирается та фонема, для которой суммарное сходство окажется большим.

Разность между двумя суммарными сходствами можно представить как сумму двух разностей сходства. Одна из них, обозначим ее как $\psi_{m, b}(x_1)$, представляет разность сходства между [m] и [b] по параметру x_1 , вторая, обозначим ее как $\psi_{m, b}(x_2)$, представляет разность сходства между теми же фонемами по параметру x_2 .

Так как [m] и [b] являются соседними фонемами, а x_1 и x_2 мы будем считать аналогами параметров I_4 и ΔF_0 , функции $\psi_{m, b}(x_1)$ и $\psi_{m, b}(x_2)$ должны иметь такой вид, как показано на рис. 1, В и В.

Рассмотрим теперь эксперимент по поиску фонемной границы, в котором экспериментатор управляет параметром x_2 , а испытуемый — параметром x_1 . Выбрав некоторое значение x_2 , экспериментатор тем самым задает определенное значение разности сходства $\psi_{m, b}(x_2)$. Испытуемый, управляя x_1 , стремится найти такое значение x_1 , при котором суммарное сходство сигнала с [m] будет равно суммарному сходству с [b], т. е.

$$\psi_{m, b}(x_1) + \psi_{m, b}(x_2) = 0.$$

Предположим, что экспериментатор выбрал значение x_2 , лежащее справа от $a_2 + \alpha_2$ (оно указано стрелкой на рис. 1, В). Так как в этом случае $\psi_{m, b}(x_2) = -k_2$, испытуемому нужно найти такое значение x_1 , при котором

$$\psi_{m, b}(x_1) - k_2 = 0.$$

Понятно, что решить эту задачу можно лишь при условии, если максимальное возможное значение $\psi_{m, b}(x_1)$ (оно обозначено k_1 на рис. 1, В) превышает k_2 или равно ему. В противном случае при всех значениях x_1 сигнал будет оставаться одной и той же фонемой, в данном случае фонемой [b].

Иными словами, определение границы по параметру x_1 возможно при любых значениях x_2 , только если x_1 является более «важным» параметром ($k_1 > k_2$). Так как из двух параметров какой-то должен, очевидно, оказаться более важным, из модели следует, что граница в $x_1 x_2$ плоскости должна разделить эту плоскость на две области, причем это деление должно проходить по более важному параметру.

Рассмотрим теперь влияние менее важного параметра (пусть это будет x_2) на положение границы. Легко заметить, что максимальное изменение положения границы по x_1 в зависимости от изменения x_2 будет наблюдаться в том случае, если k_2 почти равно k_1 , т. е. если параметры примерно равны по важности.

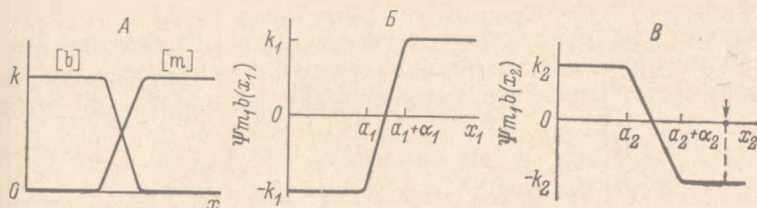


Рис. 1. Схематическая иллюстрация принятых предположений относительно вида функции сходства.

А — зависимость между значением сигнала по акустическому параметру x и величиной сходства сигнала с фонемами [b] и [m]; Б — предполагаемый вид функции разности сходства по параметру x_1 ; В — предполагаемый вид функции разности сходства по параметру x_2 . Объяснение в тексте.

При $x_2 < a_2$, что соответствует $\psi_{m,b}(x_2) = k_2$, граница будет устанавливаться на $x = a_1$, так как при этом $\psi_{m,b}(x_1) = -k_1$. При $x_2 > a_2 + \alpha_2$ граница будет устанавливаться на $x_1 = a_1 + \alpha_1$.

Таким образом, изменения значений x_2 , устанавливаемых экспериментатором, должны приводить к закономерным измене-

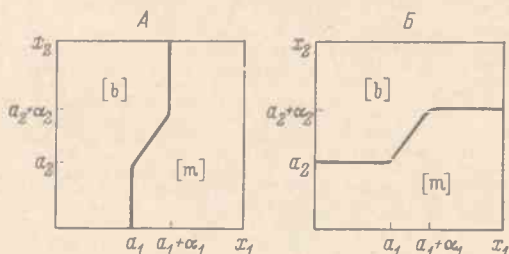


Рис. 2. Ожидаемый вид границы между фонемами [b] и [m] в x_1x_2 плоскости.

А — параметр x_1 более важен, чем параметр x_2 ;
Б — параметр x_2 более важен, чем параметр x_1 .
Объяснение в тексте.

ниям положения границы по x_1 , однако граница не может выходить за пределы, указанные a_1 и $a_1 + \alpha_1$.

На рис. 2 показано, как должна выглядеть граница между [m] и [b] в x_1x_2 плоскости, если x_1 является более важным параметром (рис. 2, А) и если x_2 является более важным (рис. 2, Б).

Результаты, полученные в эксперименте, приведены на рис. 3—5. Можно видеть, что они качественно согласуются с мо-

делью в том смысле, что область изменения положения границы по одному параметру при изменении другого параметра является достаточно узкой; направление изменения является таким, как и ожидалось. Вместе с тем тот факт, что испытуемые оказались способными определить границу по каждому параметру, является несовместимым с моделью.

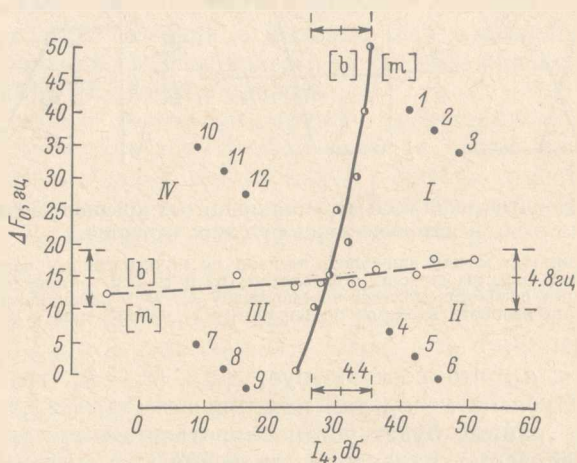


Рис. 3. Результаты экспериментального определения границы между фонемами [b] и [m] в пространстве параметров I_4 и ΔF_0 . Испытуемый П. Т.

Кружками, соединенными прерывистой линией, показаны данные первой серии опытов, когда испытуемый искал границу между [b] и [m], изменяя ΔF_0 ; каждая точка — среднее арифметическое из 20 значений установки границы (слева и справа отрезками показаны среднее значение границы и ее стандартное отклонение, вычисленные по объединенным данным (при всех исследованных значениях параметра I_4). Черно-белыми кружками, соединенными сплошной линией, показаны данные второй серии опытов, когда испытуемый искал границу между [b] и [m], изменяя I_4 ; каждая точка — среднее из 10 значений установки границы (сверху и снизу — среднее значение границы и ее стандартное отклонение по объединенным данным). Точками 1—12 обозначены стимулы, применявшиеся в экспериментах по шкалированию субъективных расстояний (объяснение в тексте).

Очевидно, что причина противоречия кроется в том, что рассмотренная модель обладала постоянными параметрами. Предполагалось не только то, что функции сходства постоянны, но и что при выборе фонемы всегда используется один и тот же критерий.

Для модели с постоянными параметрами связь между физическими характеристиками сигнала и его фонемной интерпретацией является неизменной, хотя в силу «шумов» при обработке и может иметь вероятностный характер. Результаты эксперимента, напротив того, показывают, что характер интерпретации одних и тех же сигналов совершенно закономерно изменяется

при изменении ситуации. Рассматривая границу по ΔF_0 (прерывистая линия на рис. 3—5), мы видим, что стимулы из области I интерпретируются испытуемыми как [b]. Рассматривая границу по I_4 , мы убеждаемся, что теперь эти же стимулы интерпретируются как [m].

Для того чтобы такой эффект был возможен в модели, необходимо допустить, что или функция сходства может меняться (k_1 и k_2 являются управляемыми параметрами), или может меняться критерий. Последнее предположение означает, что в уравнении

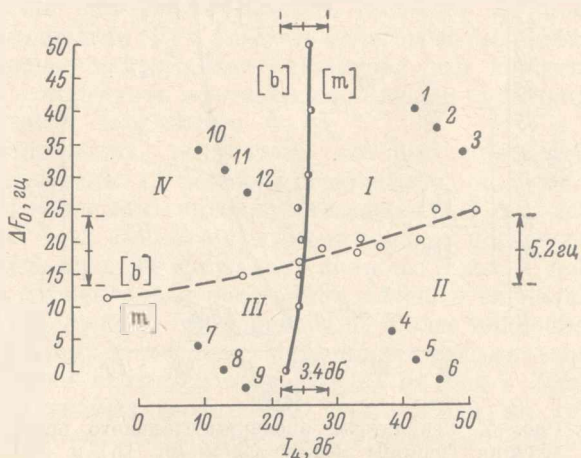


Рис. 4. Результаты экспериментального определения границы между фонемами [b] и [m] в пространстве параметров I_4 и ΔF_0 . Испытуемый Л. Ч.

Обозначения те же, что на рис. 3.

границы имеется дополнительный коэффициент d , значение которого может подбираться так, чтобы границу можно было найти, т. е. чтобы выполнялось условие:

$$\psi_{m, b}(x_1) + \psi_{m, b}(x_2) + d = 0.$$

Описание испытуемыми их субъективных впечатлений об опыте согласуется с предположением об изменении критерия. Испытуемые объясняли, что стимулы из области I похожи и на [b] и на [m]. Если они воспринимались наряду с явными [m] (поиск границы по ΔF_0), легко было настроиться на то, чтобы принимать их за [b]. Рядом с явными [b] те же стимулы из области I легко интерпретировались как [m].

Существенно подчеркнуть, что модель с переменным критерием, обеспечивая возможность определения границы по каждому из параметров, должна вместе с тем давать менее регулярный вид границы, чем ранее рассмотренная модель. Единственное,

что должно выполняться, это нахождение границы по x_1 в пределах полосы от a_1 до $a_1 + \alpha_1$, а границы по x_2 — в пределах от a_2 до $a_2 + \alpha_2$.

Полученные данные показывают, что эта полоса, где может находиться граница, является достаточно узкой — не превышает 15 дБ по I_4 и 13 гц по ΔF_0 . Вместе с тем сам факт смещения границы является достоверным. Так, влияние параметра I_4 на положение границы по ΔF_0 оказалось существенным при 1%-м уровне зна-

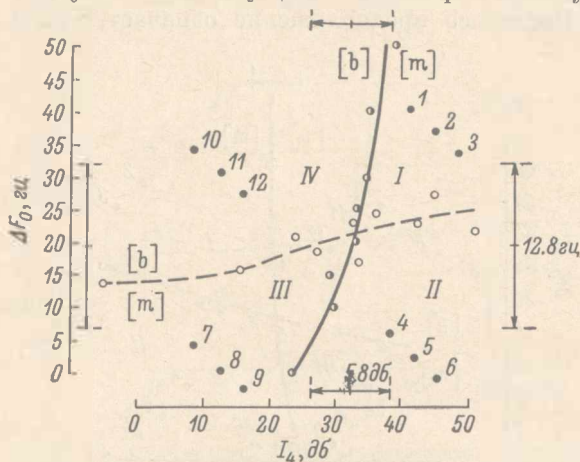


Рис. 5. Результаты экспериментального определения границы между фонемами [b] и [m] в пространстве параметров I_4 и ΔF_0 . Испытуемый Ю. К.

Обозначения те же, что на рис. 3.

чимости для испытуемых П. Т. и Л. Ч. и при 5%-м уровне значимости для испытуемого Ю. К.

Итак, данные, полученные методом поиска границы, качественно согласуются с простыми предположениями относительно функции сходства. Единственное усложнение гипотезы, требовавшееся для описания данных, состоит в допущении возможности управления критерием, применяемым при принятии решения. Однако при этом усложнении модели метод поиска границы принципиально не позволяет получить информацию об относительной важности отдельных параметров. На следующем этапе работы для этой цели использовался метод психологического шкалирования субъективных расстояний между стимулом и фонемой.

ШКАЛИРОВАНИЕ СУБЪЕКТИВНЫХ РАССТОЯНИЙ

Для сигналов, расположенных достаточно далеко от фонемных границ (рис. 3—5), можно считать, что их сходство с [b] или [m] по каждому параметру близко или к нулю, или к максимальному

значению. Обозначим максимальные значения сходства как $k_{1,b}$, $k_{1,m}$, $k_{2,b}$, $k_{2,m}$. Выпишем для далеких от границ сигналов из областей I—IV (рис. 3—5), какие значения суммарного сходства ψ_b и ψ_m должны им соответствовать.

Область I: $\psi_b = k_{2,b}$; $\psi_m = k_{1,m}$. Область II: $\psi_b = 0$; $\psi_m = k_{1,m} + k_{2,m}$.
Область III: $\psi_b = k_{1,b}$; $\psi_m = k_{2,m}$. Область IV: $\psi_b = k_{1,b} + k_{2,b}$; $\psi_m = 0$.

Отсюда видно, что если можно было бы каким-то методом экспериментально определить сходство сигналов из этих четырех областей с фонемами [b] и [m], это дало бы возможность не только найти значения максимального сходства для каждого из параметров, но и проверить допущение о простом суммировании сходства по разным параметрам.

В экспериментах, результаты которых описываются ниже, были использованы 12 различных сигналов, по 3 из каждой области. Положение этих сигналов в плоскости $I_4 \Delta F_0$ показано точками на рис. 3—5. Испытуемые, прослушивая последовательность, составленную из этих сигналов, должны были в одном случае определять субъективное расстояние каждого сигнала от фонемы [b], в другом случае — от фонемы [m] (более подробное описание эксперимента см. в разделе «Методика»). Субъективное расстояние оценивалось испытуемым в баллах от 0 до 7. Метод исследования субъективных расстояний был выбран по той причине, что он является достаточно разработанным (Torgerson, 1958; Голузина и др., 1966). Имеется опыт применения его при исследовании восприятия речевых сигналов (Галунов, 1967; Голузина, 1971, и др.).

Обработка материала была начата с построения суммарных (по всем испытуемым) гистограмм оценок расстояний для сигналов областей I—IV. Целью этой обработки была проверка гипотезы, что человек на самом деле не может оценивать сходства звука с фонемой, он может только выбрать одну или другую фонему. В этом случае следовало ожидать, что все четыре распределения оценок должны представлять собой суммы двух распределений, одно из которых соответствует выбору [m], а другое — выбору [b]. Распределение оценок для области IV должно соответствовать выбору [b], для области II — выбору [m]. Понятно, что распределения для областей I и III должны представлять собой тогда суммы взятых с разными коэффициентами распределений для областей IV и II.

Результаты, приведенные на рис. 6, показывают, что эту гипотезу необходимо отвергнуть — распределения I и III несомненно являются самостоятельными, их нельзя получить путем суммирования распределений IV и II. Таким образом, можно утверждать, что человек располагает какими-то представлениями о сходстве сигнала с фонемой.

В целях упрощения дальнейшего рассмотрения данных субъективное расстояние было превращено в субъективное сходство, которое определялось как разность между максимальным возможным баллом (7) и баллом, названным при оценке расстояния.

Дальше это будет называться субъективной оценкой сходства или просто оценкой сходства. Среднее значение оценки сходства (в баллах) будет обозначаться буквой A с индексами, указывающими фонему и область

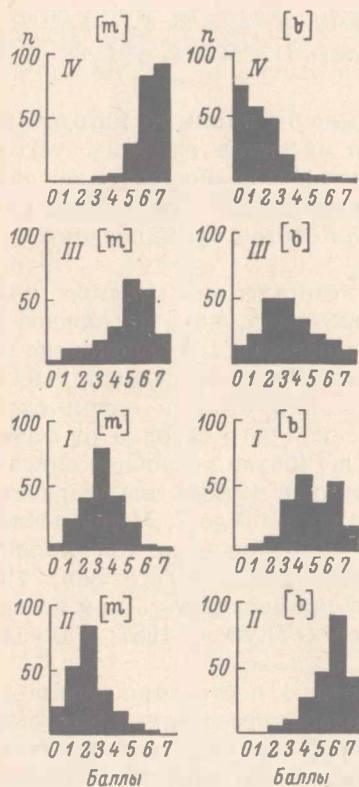


Рис. 6. Суммарные (по всем испытуемым) гистограммы оценок субъективного расстояния сигнала от фонемы [b] и от фонемы [m] для стимулов областей IV, III, I, II.

По оси абсцисс — оценка расстояния, баллы; по оси ординат — число оценок.

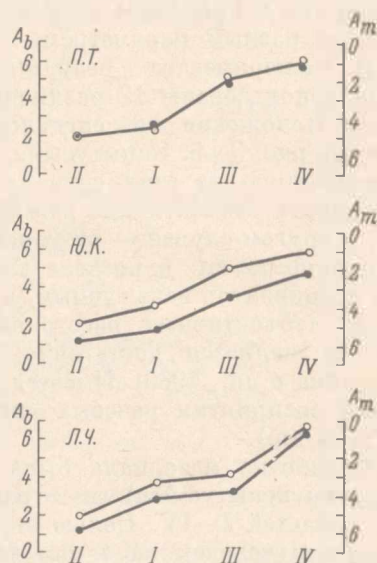


Рис. 7. Средние значения оценки сходства с фонемой [b] (точки) и с фонемой [m] (кружки) для стимулов областей I, II, III, IV. Данные разных испытуемых.

Левая шкала указывает сходство с фонемой [b], правая — с фонемой [m].

сигналов. Например, $A_{b,III}$ — среднее значение оценки сходства сигналов из области III с фонемой [b].

На рис. 7 приведены значения A_b и A_m для сигналов разных областей. Можно видеть, что сигналы образуют один и тот же ряд (II, I, III, IV) по возрастанию A_b и убыванию A_m . Порядок сигналов в ряду оказывается одним и тем же у всех трех испытуемых.

Считая, что субъективная оценка сходства является возрастающей функцией суммарного сходства звука с фонемой, из полученных данных следует заключить, что $k_{1,b} > k_{2,b}$ и $k_{1,m} > k_{2,m}$, т. е. что параметр I_4 является более важным, чем параметр ΔF_0 .

Если допустить, что субъективная оценка сходства связана с суммарным сходством линейной зависимостью, по полученным значениям оценок сходства можно вычислить отношения $k_{2,b}/k_{1,b}$ и $k_{2,m}/k_{1,m}$:

$$\frac{k_{2,b}}{k_{1,b}} = \frac{A_{b,I} - A_{b,II}}{A_{b,III} - A_{b,II}};$$

$$\frac{k_{2,m}}{k_{1,m}} = \frac{A_{m,III} - A_{m,IV}}{A_{m,I} - A_{m,IV}}.$$

Вычисленные таким образом значения $k_{2,b}/k_{1,b}$ и $k_{2,m}/k_{1,m}$ оказались следующими: 0.22 и 0.21 (испытуемый П. Т.); 0.22 и 0.30 (испытуемый Ю. К.); 0.88 и 0.89 (испытуемый Л. Ч.). Можно видеть, что значения $k_{2,b}/k_{1,b}$ и $k_{2,m}/k_{1,m}$ практически совпадают. Для испытуемых П. Т. и Ю. К. параметр ΔF_0 оказывается значительно менее важным, чем параметр I_4 , для испытуемого Л. Ч. различие в важности параметров мало.

При предположении линейной зависимости между оценкой сходства и суммарным сходством можно также проверить гипотезу о простом суммировании значений сходства по отдельным параметрам. Если эта гипотеза верна, то

$$A_{b,I} + A_{b,III} = A_{b,II} + A_{b,IV} \text{ и } A_{m,I} + A_{m,III} = A_{m,II} + A_{m,IV}.$$

Вычисления показали, что различие между этими величинами статистически недостоверно (применялся критерий Стьюдента по группированным выборкам). Следовательно, экспериментальные

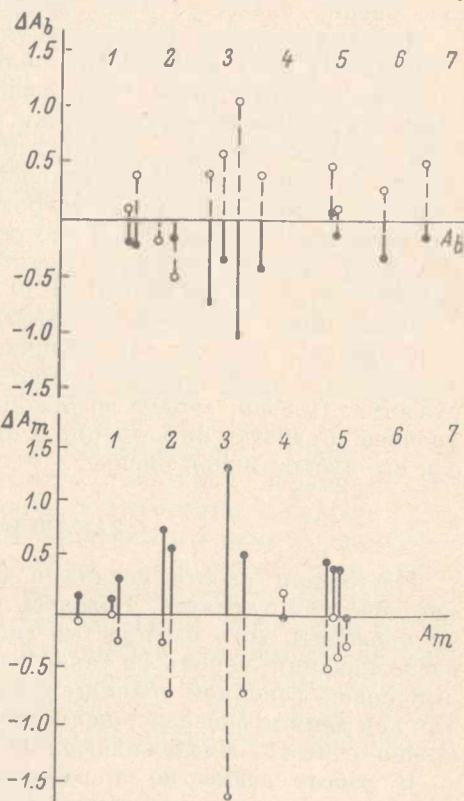


Рис. 8. Изменение оценки сходства под влиянием предшествующего стимула.

По оси абсцисс — среднее значение оценки сходства; по оси ординат — отклонение от среднего для случаев, когда предшествующий стимул относился к области II (точки) и когда он относился к области IV (кружки). Верхний график — оценка сходства с фонемой [b] нижний — с фонемой [m].

данные не противоречат гипотезе о простом суммировании значений сходства по отдельным параметрам.

Последней задачей обработки материала являлась проверка наличия или отсутствия контекстуальных влияний на субъективную оценку сходства. Были раздельно вычислены значения A_b (и A_m) для случаев, когда стимулы следовали за явными [m] (область II) и за явными [b] (область IV). Далее определялась величина сдвига оценки, т. е. разность между полученной величиной и средним значением оценки, вычисленным без учета характера предшествующего стимула. Полученные данные приведены на рис. 8. Можно видеть, что сходство с [b] растет, если стимулы следуют за явным [b], после [m] оно, наоборот, падает. То же самое касается сходства с [m]: оно растет после [m] и падает после [b]. Этот эффект, носящий характер ассимиляции, наблюдался, как можно видеть на рис. 8, в 40 случаях из 48.

Дальнейшее исследование эффекта представляет очень большой интерес, так как его наличие может свидетельствовать об ошибочности принятой выше гипотезы о неизменяемости функции сходства. Правда, вполне возможно, что эффект касается высших уровней обработки информации и относится не к самому сходству, а к его субъективной оценке.

ЗАКЛЮЧЕНИЕ

Настоящая работа носила в основном разведывательный и методический характер. Казалось полезным выяснить, возможно ли проверять хотя бы простые гипотезы о процедуре фонемного распознавания, используя такие методы исследования, как активный поиск фонемной границы и психологическое шкалирование. Так как данные оказались непротиворечивыми, результаты работы можно считать обнадеживающими.

В работе совместно проверялось два предположения, одно из которых может иметь общий характер (суммирование значений сходства по отдельным параметрам), а второе, касающееся специфического вида функции сходства, является частным, оно явно не распространяется на случай гласных.

Дальнейшую проверку гипотезы о суммировании сходства целесообразно провести на стимулах с большим числом полезных параметров и с расширением набора опознаваемых фонем. Вероятно, наиболее удобным для исследования примером окажутся синтетические согласные (слоги).

Прежде чем исследовать вид функции сходства для гласных, важно проверить справедливость допущения о неизменяемости функции сходства у человека, уже обученного языку. Именно для гласных известны значительные контекстуальные влияния при восприятии (Fry et al., 1962; Lindner, 1966; Hiki et al., 1968), несомненно также, что при распознавании гласных должна произ-

водиться какая-то подстройка под диктора. Существенно выяснить, возможно ли свести эту текущую перестройку поведения к регулированию порогов в блоке принятия решений или же ее нельзя объяснить иначе как изменением функции сходства.

Результаты, полученные в настоящей работе, могут рассматриваться в связи с широко обсуждаемым в литературе представлением о категориальном характере восприятия согласных. В зависимости от того, какой смысл вкладывается в понятие категориальности, полученные данные или противоречат этому представлению, или согласуются с ним. Одно возможное толкование категориальности состоит в том, что информация, полученная испытуемым при восприятии согласного стимула, исчерпывается знанием фонемы, выбранной в ответ на стимул (Liberman et al., 1957). С этим толкованием данные не согласуются, так как испытуемые, без сомнения, располагают большей информацией — они способны указать величину сходства стимула с той и другой фонемой.

Другое возможное толкование состоит в том, что функция сходства с фонемой по отдельному параметру имеет вид, близкий к прямоугольному — интервал значений параметра, на котором сходство изменяется от нуля до максимального значения, существенно уже области допустимых значений параметра. Это толкование вполне совместимо с полученными данными.

Следует заметить, что для экспериментального разделения этих двух возможных представлений о категориальности нужны были эксперименты со стимулами, изменяющимися по двум или более параметрам. Вывод о категориальности восприятия согласных был сделан сотрудниками Хаскинских лабораторий на основании экспериментов по АВХ различению одномерных стимулов (Liberman et al., 1957, 1961). Полученные в этих экспериментах результаты совместимы с обоими толкованиями категориальности.

Л и т е р а т у р а

- Г а л у н о в В. И. 1967. Структура множества речевых образов. Дисс. Л.
Г о л у з и н а А. Г. 1971. Шкалирование субъективных расстояний между синтетическими гласными. Наст. сб.: 155—160.
Г о л у з и н а А. Г., В. В. Л ю б л и н с к а я, М. К. Р о х т л а, Л. А. Ч и с т о в и ч. 1966. Шкалирование субъективных расстояний как метод исследования слухового восприятия. В сб.: Механизмы речеобразования и восприятия сложных звуков, М.—Л.: 96—110.
Ч и с т о в и ч Л. А. 1968. Изменение основной частоты голоса как различительный признак согласных. Акуст. журн., 14: 449—456.
F a n t G. 1960. Acoustic theory of speech production. 's-Gravenhage.
F r y D. B., A. S. A b r a m s o n, P. D. E i m a s, A. M. L i b e r m a n. 1962. The identification and discrimination of synthetic vowels. Language a. Speech, 5: 171—189.
H i k i Sh., H. S a t o, T. I g a r a s h i, Ju. O i z u m i. 1968. Dynamic model of vowel perception. Reports of 6th Intern. Congr. on Acoustics, Tokyo, 2, В: 87—90.

- Kacprowski J., W. Mikiel. 1966. Simplified rules for parametric synthesis of nasal and stop consonants in CV syllables by means of the «terminal-analog» speech synthesizer. *Acustica*, 18 : 356—364.
- Ladefoged P., D. E. Broadbent. 1957. Information conveyed by vowels. *J. Acoust. Soc. Amer.*, 29 : 98—104.
- Lieberman A. M., K. S. Harris, H. S. Hoffman, B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *J. Exper. Psychol.*, 54 : 358—368.
- Lieberman A. M., K. S. Harris, J. A. Kinney, H. Lane. 1961. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. Exper. Psychol.*, 61 : 379—388.
- Lindner G. 1966. Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes. *Ztschr. f. Phonetik*, 19 : 287—307.
- Torgerson W. S. 1958. Theory and method of scaling. N. Y.

О ПОЛОЖЕНИИ ФОНЕМНОЙ ГРАНИЦЫ МЕЖДУ СИНТЕТИЧЕСКИМИ ГЛАСНЫМИ [i]—[e], [ü]—[ö] И [u]—[o]

Н. А. Слепокурова

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

В экспериментах В. Н. Мушникова и Л. А. Чистович (1971) с синтетическими двухформантными гласными было получено, что граница между [i] и [e], определенная методом активного поиска, имеет постоянное значение по F_1 в широком диапазоне значений частоты второй форманты стимулов.

Казалось заманчивым предположить, что та же самая граница разделяет между собой фонемы [ü] и [ö], а также [u] и [o].

Имеющиеся в литературе данные по идентификации большого набора синтетических гласных (Janota, 1967; Scholes, 1967; Tomisawa et al., 1968, и др.) не дают оснований ни принять, ни отвергнуть эту гипотезу. Контуры фонемных областей в плоскости F_1F_2 , приводимые авторами, имеют весьма причудливый вид и сильно отличаются не только в разных работах, но и у разных групп испытуемых в одной и той же работе.

В кратком сообщении Чистович и др. (Chistovich et al., 1966) имеются указания на то, что одна и та же граница может разделять [i] и [e], [ü] и [ö], [u] и [o], однако авторы подчеркивают, что полученные данные имеют сугубо ориентировочный характер.

В настоящей работе была сделана попытка экспериментально проверить гипотезу о постоянстве границ для указанных трех пар гласных, пользуясь методикой активного поиска границы.

МЕТОДИКА

В качестве экспериментального материала использовались синтетические двухформантные гласные, имеющие длительность 200 мсек. и частоту основного тона 115 гц. Непрерывная последовательность гласных непосредственно с синтезатора подавалась на телефоны испытуемому. Экспериментатор устанавливал фиксированное значение F_2 , которое оставалось постоянным в течение приблизительно 20 поисков. Управление F_1 производилось с помощью двух параллельно включенных магазинов емкостей. Значение на одном магазине устанавливалось экспериментатором, испытуемый управлял вторым магазином. Было принято, что испытуемый мог производить поиск границы только в одном направлении. После того как испытуемый устанавливал границу, экспериментатор выставлял на магазине новое значение емкости, и поиск начинался сначала. Эксперимент производился мелкими сериями, в каждой из которых испытуемый при данном значении параметров производил 5—10 установок границы. Частоты второй форманты подбирались таким образом, чтобы исходный гласный (гласный, от которого начинался поиск) звучал как можно более естественно. Такими значениями F_2 были 800 гц для гласных [u]—[o], 1400 гц для гласных [ü]—[ö] и 2200 гц для гласных [i]—[e].

В эксперименте приняли участие 8 испытуемых.

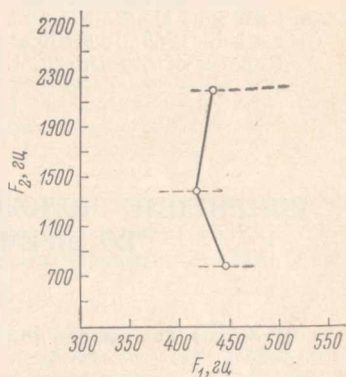
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Прежде всего оказалось, что только 5 из 8 испытуемых смогли определить границу между [ü] и [ö]. Очевидно, это объясняется отсутствием в русском языке выше-названных гласных как самостоятельных лингвистических фонем.

Далее обнаружилось, что разброс результатов является весьма большим не только для разных испытуемых, но и между средними у одного испытуемого в повторных сериях из 10 измерений. Различие между средними по сериям для одной и той же пары гласных достигало 50 гц. Максимальная разница между испытуемыми достигала 50 гц.

Этот факт позволяет использовать только средние по всей группе. Они приведены на рисунке.

Из графика видно, что разница между высшей и низшей точками средней границы сравнительно мала. Достоверность различий проверялась по критерию Стьюдента. Оказалось, что различия недостоверны, но этот результат является сомнительным, так как распределение оценок, полученных в результате эксперимента, не является нормальным.



Средняя граница по F_1 (ось абсцисс) для пар гласных [i]—[e], [ü]—[ö] и [u]—[o] при меняющихся значениях F_2 (ось ординат).

Горизонтальные штриховые линии показывают диапазон разброса границ для отдельных испытуемых.

Проведенное Клещевым (1971) рассмотрение данных, полученных Мушниковым и Чистович при определении границы между [i] и [e], также говорит о том, что гипотезу о случайных флюктуациях границ нужно отвергнуть. Закономерные сдвиги границ были показаны и в работе Кузьмина и Лисенко (1971).

Таким образом, на основании полученного материала пока можно только утверждать, что гипотеза о постоянстве границ не отвергается. Для более надежных выводов необходима дополнительная работа по изучению закономерностей сдвига границ в эксперименте и разработка метода, позволяющего получать более стабильные результаты.

Л и т е р а т у р а

- К л е щ е в А. С. 1971. Изменение положения фонемной границы во время эксперимента. Наст. сб. : 138—142.
- К у з ь м и н Ю. И., Д. М. Л и с е н к о. 1971. Контекстуальные влияния в экспериментах по определению фонемных границ. Наст. сб. : 142—149.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971. О слуховом описании гласного. I. Признаки, различающие [i] и [e]. Наст. сб. : 5—10.
- C h i s t o v i c h L., G. F a n t, A. d e S e r p a - L e i t a o. 1966. Miming and perception of synthetic vowels, part II. STL—QPSR, № 3.
- J a n o t a P. 1967. Personal characteristics of speech. Praha.
- S c h o l e s R. J. 1967. Categorical responses to synthetic vocalic stimuli by speakers of various languages. Language a. Speech, 10, 252—282.
- T o m i s a w a M., I. K i r i k a e, T. S h i t a r a, T. S a t o, H. F u j i - s a k i. 1968. Identification of synthetic vowels in perceptive deafness. Reports of 6th Intern. Congr. on Acoustics, Tokyo, A : 77—80.

ИЗМЕНЕНИЕ ПОЛОЖЕНИЯ ФОНЕМНОЙ ГРАНИЦЫ ВО ВРЕМЯ ЭКСПЕРИМЕНТА

А. С. Клещев

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

Понятие фонемной границы возникло из опытов по идентификации фонем. В этих экспериментах за фонемную границу условно принималась та точка на оси некоторого акустического параметра, в которой эмпирическая функция идентификации равнялась $\frac{1}{2}$. Такое условное понятие позволяло применять простую модель идентификации, согласно которой испытуемый принимает то или иное решение о принадлежности стимула к фонемной категории в зависимости от того, по какую сторону от фонемной границы оказался на оси предъявленный стимул. При этом предполагалось, что фонемная граница подвержена случайным флюктуациям и имеет распределение, близкое к нормальному.

С фонемной границей связаны некоторые эффекты, полученные в экспериментах по идентификации и различению речевых звуков. Оказалось, что в окрестности фонемной границы наблюдается увеличение частоты различения стимулов (Liberman et al., 1957), а также увеличение времени, необходимого для идентификации (Studdert-Kennedy et al., 1963). Поскольку свойства фонемных границ должны быть тесно связаны с видом правил фонемной классификации, изучение этих свойств представляет определенный интерес.

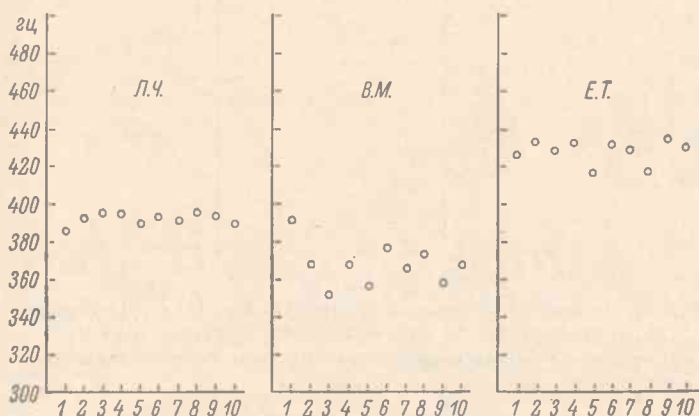


Рис. 1. Положение фонемной границы между гласными [i] и [e] на оси частот первой форманты (по оси ординат) при ряде последовательных измерений границы (по оси абсцисс — номер измерения). Данные трех испытуемых.

Ниже приведены некоторые факты, полученные в экспериментах по определению положения фонемной границы между гласными [i] и [e] методом активного поиска (Мушников, Чистович, 1971).

В качестве акустического параметра в этом эксперименте была выбрана частота первой форманты. В эксперименте участвовало трое испытуемых. Весь эксперимент был разделен на серии по 10 измерений фонемной границы в каждой. Количество серий, проведенных с каждым испытуемым, было: Л. Ч. — 16, В. М. — 15, Е. Т. — 16.

Исходя из приведенной выше простой модели, естественно было ожидать, что этот метод должен давать то же положение фонемной границы, что и опыты по идентификации и различению. Суммарное распределение результатов, как и распределения в отдельных сериях, должны быть близки к нормальному. Разброс результатов должен быть велик и складываться из разброса, вызванного флюктуацией границы, и разброса, связанного с принятием решения (остановкой поиска).

Измерения, полученные в разных сериях эксперимента, можно условно разделить на три приблизительно равные по количеству группы. К первой группе (рис. 1) относятся серии, в которых указанные положения границы группируются вблизи некоторого

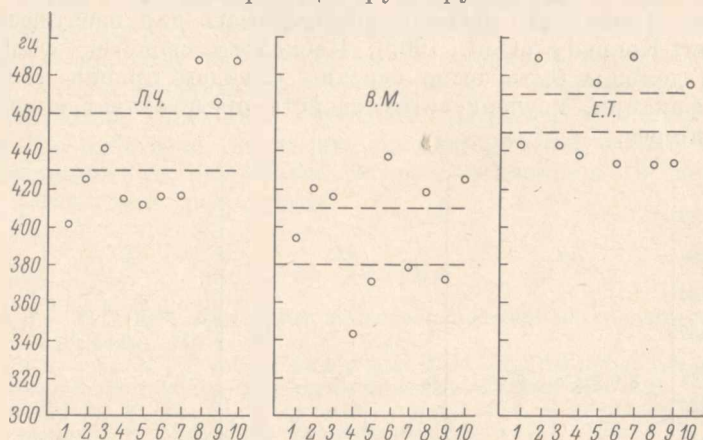


Рис. 2. Положение фонемной границы между [i] и [e] в сериях, характеризующихся преобладанием крайних значений. Прерывистые линии ограничивают среднюю область значений.

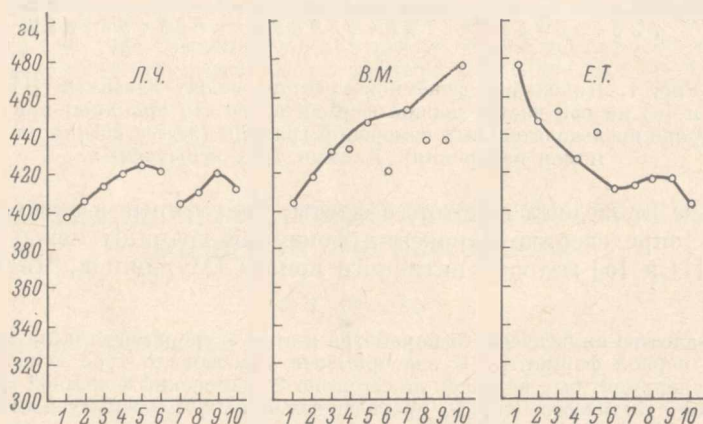


Рис. 3. Положение фонемной границы между [i] и [e] в сериях, характеризующихся монотонным смещением границы.

постоянного значения. Разброс результатов в сериях этой группы весьма незначителен и, по-видимому, связан с остановкой поиска. Эта группа измерений не подтверждает предположения о флюктуации фонемной границы.

Серии, отнесенные ко второй группе, иллюстрируются рис. 2. Измерения каждой такой серии группируются около концов диапа-

зона значений в серии, в то время как частота попаданий в среднюю область (ограниченную на рисунке прерывистыми линиями) очень мала. Эта группа опровергает предположение о нормальности распределения измерений в сериях.

Наконец, рис. 3 представляет серии, отнесенные к третьей группе. Эти серии отличаются тем, что в них можно выделить участки монотонного (или почти монотонного) изменения измеряемой величины в зависимости от номера наблюдения в серии.

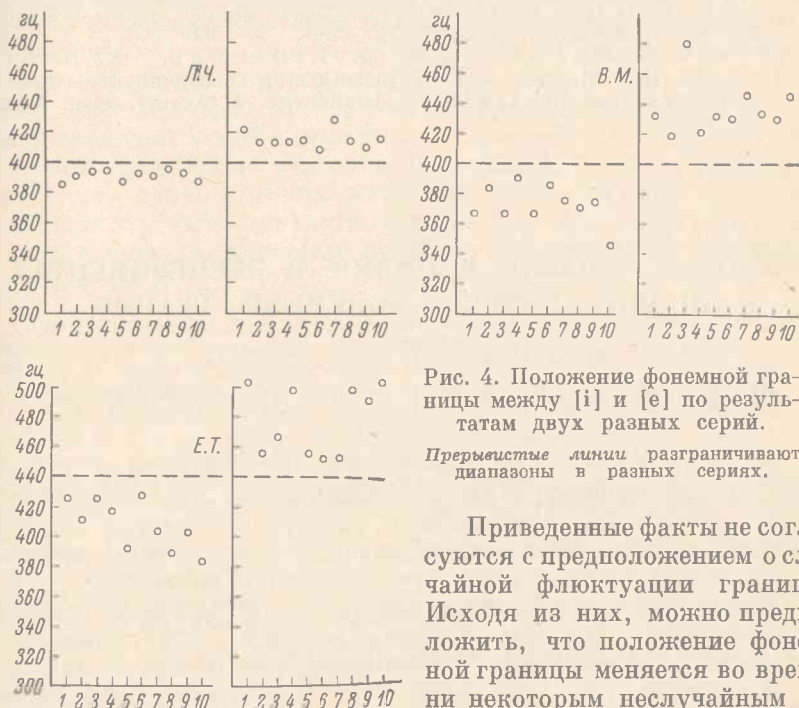


Рис. 4. Положение фонемной границы между [i] и [e] по результатам двух разных серий.

Прерывистые линии разграничивают диапазоны в разных сериях.

Приведенные факты не согласуются с предположением о случайной флюктуации границы. Исходя из них, можно предположить, что положение фонемной границы меняется во времени некоторым неслучайным образом. Это предположение подтверждается и следующим интересным фактом, иллюстрируемым рис. 4. На этом рисунке представлены пары серий, причем обе серии в паре принадлежат одному и тому же испытуемому. Как видно из рисунка, диапазоны значений в разных сериях каждой пары не перекрываются (граница между диапазонами проведена прерывистой линией). Этого достаточно, чтобы утверждать (с крайне малой вероятностью ошибки), что измерения в этих сериях принадлежат к разным распределениям. Иными словами, положение фонемной границы в этих сериях было существенно различным.

Таким образом, результаты эксперимента по активному поиску фонемной границы показывают, что изменение положения фонемной границы является некоторым неслучайным процессом. Выяс-

нение причин, вызывающих такие изменения, а также изучение закономерностей процесса изменения положения фонемной границы могут стать предметом интересных исследований.

Литература

- Мушников В. Н., Л. А. Чистович. 1971. О слуховом описании гласного. I. Признаки, различающие [i] и [e]. Наст. сб.: 5—10.
- Lieberman A. M., K. S. Harris, H. S. Hoffman, B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exper. Psychol., 54: 358—368.
- Studdert-Kennedy M., A. M. Liberman, K. N. Stevens. 1963. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. J. Acoust. Soc. Amer., 35: 1009.

КОНТЕКСТУАЛЬНЫЕ ВЛИЯНИЯ В ЭКСПЕРИМЕНТАХ ПО ОПРЕДЕЛЕНИЮ ФОНЕМНЫХ ГРАНИЦ

Ю. И. Кузьмин, Д. М. Лисенко

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

Одним из существенных требований, предъявляемых к модели, описывающей процедуру распознавания фонем, является способность к адаптации, подстройке к диктору. В частности, такая способность предполагает текущее изменение фонемной границы в зависимости от ансамбля предъявляемых сигналов.¹

Способность человека подстраиваться к диктору была показана в нескольких работах (Ladefoged, Broadbent, 1957; Broadbent, Ladefoged, 1960; Fujisaki, Kawashima, 1967). В первых двух работах исследовалось восприятие слов [bi:t], [bit], [bet], [bæt] в окружении различных ключевых фраз. Все ключевые фразы были получены из одной естественной путем изменения частот формант. Восприятие слова закономерно зависело от того окружения, в котором оно было предъявлено. В работе Фуджисаки и Кавашима был показан существенный сдвиг фонемных границ в плоскости F_1 — F_2 в зависимости от частоты основного тона и частот верхних формант стимулов.

Аналогичные явления наблюдались и в экспериментах с активным поиском границ между синтетическими гласными (Муш-

¹ При этом понятие фонемной границы определим следующим образом. Если имеется функция фонемной идентификации стимулов, отличающихся по одному параметру, то фонемной границей будем называть то значение стимула по данному параметру, при котором ответы слушателей будут равновероятны.

ников, Чистович, 1971). Испытуемые на протяжении ряда измерений могли определять границу с очень незначительным разбросом данных; в другой серии экспериментов те же испытуемые так же стабильно помещали границу в другом месте. То же наблюдалось нами при определении границ между стимулами, отличающимися по величине перепада интенсивности.

Однако о механизмах такой подстройки ничего не известно. Прежде всего нас интересовал вопрос, является ли такая подстройка следствием обучения или результатом действия какого-то автоматического механизма. Чтобы убедиться в возможности изменить функцию идентификации в результате некоторого обучения, был проведен эксперимент с идентификацией стимулов, представлявших собой синтетический гласный [a], на протяжении которого было резкое изменение интенсивности. Как было показано ранее, такие стимулы воспринимаются либо как слоги, содержащие [l], либо как слоги, содержащие назальный согласный, в зависимости от величины перепада интенсивности (Кузьмин, Лисенко, 1971).

Были составлены две тестовые таблицы стимулов, каждая из которых включала 8 разных сигналов, отличающихся по величине перепада интенсивности. В наборе сигналов, использованных при построении первой таблицы, величина перепада интенсивности изменялась от 4 до 22.4 дБ ступенями по 2.63 дБ. Во второй таблице величина перепада интенсивности изменялась от 12.7 до 31.1 дБ ступенями той же величины. Таким образом, стимулы первой таблицы располагались приблизительно симметрично относительно границы между согласным [l] и назальными согласными, в то время как стимулы второй таблицы лежали преимущественно в области назальных согласных (Кузьмин, Лисенко, 1971, рис. 4).

Перед началом каждой таблицы следовали сначала 10 стимулов с наименьшим перепадом интенсивности, а потом 10 стимулов с наибольшим для данной таблицы перепадом интенсивности (будем называть эти стимулы стандартными). Далее следовало 5 порций таблицы, каждая из которых включала 24 тестовых сигнала, расположенных в случайном порядке. Перед каждой из этих порций опять повторялись крайние представители использованного набора сигналов (стандарты).

Экспериментальные таблицы предъявлялись слушателям в звукозаглушенной комнате через громкоговорители. Слушателям объяснялась структура таблицы, как это изложено выше, и предлагалось реагировать на каждый стимул поворотом ключа. В ответ на первые десять стимулов они должны были поворачивать ключ влево, в ответ на следующие десять стимулов — вправо. В отношении каждого из 24 тестовых стимулов испытуемые должны были решить, на какой из стандартов он больше похож, и повернуть ключ в соответствующую сторону. Полученные при замыкании ключа электрические импульсы разной полярности записывались параллельно для всех испытуемых на многоканальном чернильном самописце вместе с огибающей предъявленного стимула.

В указанных экспериментальных условиях группа из 5 слушателей сначала дважды прослушала первую таблицу, а потом дважды прослушала вторую таблицу. В результате от каждого испытуемого было получено по 30 реакций на каждый экспериментальный стимул.

Результаты обработки полученного материала показали следующее. Два слушателя из пяти почти совершенно не справились

с заданием и реагировали на все сигналы практически случайно. Результаты остальных трех слушателей приведены на рис. 1.

Можно видеть, что испытуемые реагируют на один и тот же стимул различно в зависимости от того, в какой таблице он предъявляется. При этом порог, соответствующий 50% правых и 50% левых реакций, располагается при прослушивании обеих таблиц приблизительно в середине диапазона изменения сигнала. Так как результаты обоих прослушиваний каждой из таблиц практически совпадают, следует думать, что смещение порога происходит практически сразу после прослушивания предъявляемых

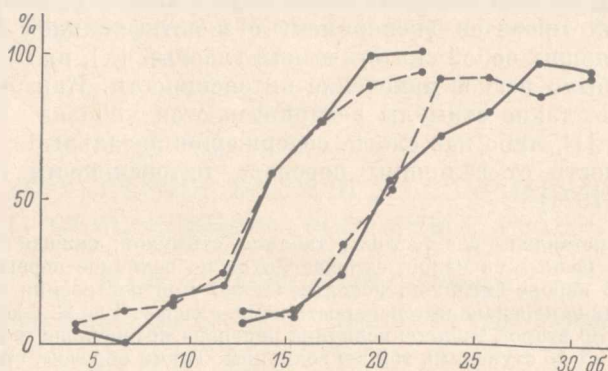


Рис. 1. Результаты классификации стимулов с резким изменением интенсивности (I эксперимент).

По оси абсцисс — величина перепада интенсивности, дБ; по оси ординат — процент «правых» реакций. Прерывистые линии — результаты первых прослушиваний таблиц, сплошные — вторых прослушиваний таблиц.

в начале таблицы стандартов. В следующем эксперименте те же самые таблицы были предъявлены другой группе слушателей при некоторой модификации инструкции. А именно, слушателям было указано, что левый стандарт в каждой таблице является слогом [lu], а правый стандарт — слогом [mu]. На [lu] следовало поворачивать ключ влево, на [mu] — вправо. В остальном экспериментальная ситуация оставалась той же самой. Из 4 слушателей, участвовавших в этом эксперименте, двое участвовали в предыдущем.

Как показали результаты, в данном случае смещение кривых наблюдалось у всех слушателей. Суммарные данные для 4 слушателей приведены на рис. 2. Прерывистыми линиями отмечены данные, полученные при первом предъявлении таблиц, сплошными линиями — при повторном предъявлении таблиц. Расхождение этих кривых в точках, обозначенных кружками, оказалось достоверным при уровне значимости 0.05. Достоверность проверялась по методу χ^2 .

Таким образом, в основном результат оказался таким же, как и в первом эксперименте. Единственное отличие заключалось в некотором смещении кривых при повторном прослушивании той же таблицы. Это заставляет думать, что обучение в данном случае происходило медленнее и не заканчивалось на протяжении однократного прослушивания таблицы.

Результаты этих экспериментов свидетельствуют о том, что критерий, применяемый слушателями при классификации использованных сигналов, не остается в данных условиях одним и тем же. Однако это еще не доказывает предполагаемой подвижности фонемных границ, хотя и делает такое предположение более вероятным. Действительно, полученный эффект мог бы наблюдаться, если бы испытуемые вопреки инструкции вообще не принимали фонемных решений и осуществляли бы выбор той или другой реакции на основании правила, вырабатываемого ими после прослушивания предъявляемых в начале эксперимента эталонов. Поэтому в дальнейшем в качестве экспериментальных стимулов были использованы довольно естественно звучащие синтетические гласные, а сама экспериментальная процедура изменена так, чтобы исключить возможность предварительного обучения.

Использованные в следующем эксперименте стимулы представляли собой синтезированные на OVE-II гласные, частоты трех первых формант которых в гц приведены ниже:

№ гласного	F_1	F_2	F_3
1	270	2500	3400
2	320	2450	3150
3	375	2420	3000
4	430	2350	2820
5	480	2250	2750
6	540	2140	2650
7	605	1950	2600
8	670	1750	2610

Эти гласные были исследованы ранее в работе Чистович и др. (Chistovich et al., 1966), а также в работе Голузиной (1971). Было показано, что первые из этих стимулов воспринимаются как гласный [i], последние — как гласный [e]. Граничное значение между звуками [i] и [e] находится, по данным ряда авторов, в районе 390—400 гц (Мушников, Чистович, 1971). Из этих стимулов были составлены две таблицы. Первая из них представляла случайную последовательность шести первых синтетических гласных, вторая — случайную последовательность последних шести гласных. (Гласные, обозначенные номерами 3, 4, 5 и 6, входили в обе таблицы). Каждый из гласных, включенных в таблицу, повторялся 20 раз в случайном порядке.

Экспериментальная ситуация в основном оставалась такой же, как и в предыдущих экспериментах. Однако, чтобы исключить возможность предварительного обучения, никакой демонстрации стимулов перед началом

эксперимента не производилось. Инструкция, даваемая слушателям, сводилась к следующему: «Вам будет предъявлена таблица синтетических гласных. Услышав гласный [i], поверните ключ налево, услышав гласный [e] — направо». В экспериментах участвовало 3 группы испытуемых по 5 человек в каждой. Первая группа дважды прослушала первую таблицу, вторая группа дважды прослушала вторую таблицу, третья группа дважды прослушала обе таблицы (сначала вторую, а потом — первую).

Полученные данные показали следующее. При прослушивании первой таблицы кривые идентификации, полученные для каждого из 10 слушателей, практически совпали. Границы, соответствующие 50% восприятия [e], оказались расположенными у всех слушателей в районе 400 гц. Различия между результатами первого

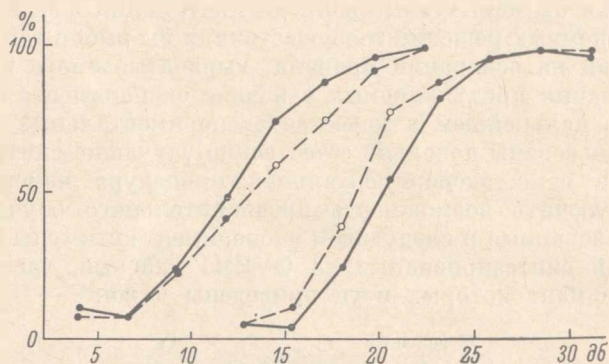


Рис. 2. Результаты классификации стимулов с резким изменением интенсивности (II эксперимент).

По оси абсцисс — величина перепада интенсивности; по оси ординат — процент «правых», или [та]-реакций. Прерывистые линии — по данным первого прослушивания таблиц, сплошные — по данным второго прослушивания.

и второго прослушивания таблицы также оказались незначительными. Это позволило объединить результаты обоих прослушиваний таблицы и высчитать средние в пределах каждой из двух групп испытуемых, работавших с этой таблицей. Полученные таким образом кривые приведены на рис. 3, А слева.

При прослушивании второй таблицы кривая идентификации осталась на том же месте только у одного из 10 слушателей. У остальных 9 слушателей кривые идентификации сдвинулись в сторону значительно больших значений частоты первой форманты. Результаты усреднения данных этих слушателей в пределах каждой из двух групп приведены на рис. 3, А справа. Можно видеть, что и в этом случае кривые довольно хорошо совпадают. Значение F_1 , соответствующее 50% восприятия [e], оказывается для обеих групп равным приблизительно 490 гц. Иными словами, граница между [i] и [e] в данном случае оказывается сдвинутой

относительно границы, полученной на предыдущем наборе сигналов, на величину порядка 100 гц.

По литературным данным, одним из показателей положения границы между двумя категориями стимулов является существенное увеличение задержки реакции при идентификации лежащих на границе сигналов (Cross, Lane, 1962; Studdert-Kennedy et al., 1963). Поэтому для той группы испытуемых, которая участвовала в прослушивании обеих таблиц, были высчитаны задержки реакций

на каждый из экспериментальных стимулов в каждой из двух экспериментальных ситуаций.

Полученные результаты приведены на рис. 3, Б. Длительно-

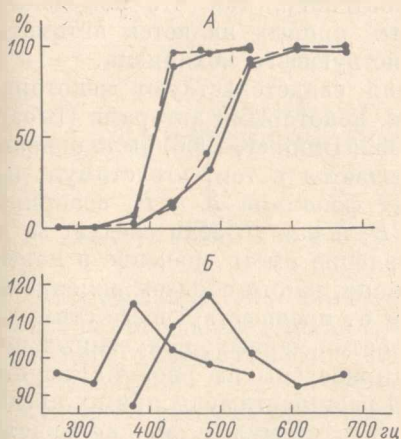


Рис. 3. Результаты классификации синтетических гласных.

По оси абсцисс — частота первой форманты стимула; по оси ординат: на А — процент [е]-ответов, на Б — длительность латентного периода реакции, % от средней. *Левые кривые* — по данным прослушивания первой таблицы, *правые* — по данным прослушивания второй таблицы. *Прерывистые линии* — при первом прослушивании таблиц, *сплошные* — при втором прослушивании.

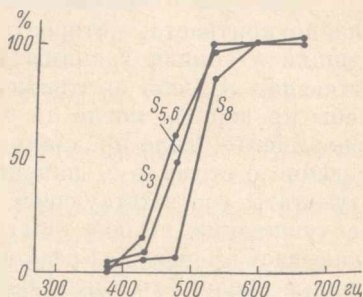


Рис. 4. Кривые идентификации гласных при их следовании после разных стимулов.

По оси абсцисс — частота F_1 ; по оси ординат — процент [е]-реакций. Параметр кривых — номер предшествующего стимула.

таблицы. Соответствующие средние оказались равными 490 мсек., при прослушивании первой таблицы и 610 мсек. — при прослушивании второй таблицы (разница в средних, по всей видимости, может быть отнесена за счет тренировки). Как показывает рисунок, сдвиг кривых идентификации сопровождается соответствующим сдвигом максимума длительности задержки.

Таким образом, смещение границы между двумя категориями стимулов может быть получено как на стимулах, принадлежность которых к речевым можно оспаривать, так и на собственно речевых стимулах. Это смещение границы сопровождается соответствующим смещением максимума длительности латентного периода. Как показывают данные, этот эффект может быть получен

сти задержек на этом графике выражены в процентах от средней для всех сигналов данной

в результате простого изменения состава экспериментальной последовательности стимулов без какой-либо предшествующей тренировки испытуемых.

Все это позволяет думать, что критерий, применяемый слушателем в процессе принятия решения о фонемах, не остается одним и тем же.

Отсутствие процесса установления в ходе эксперимента по идентификации при отсутствии обучающей последовательности свидетельствует о том, что процедура принятия фонемных решений сама по себе чувствительна к контексту, так что полученное нами смещение идентификационных кривых является автоматическим результатом постоянно действующего механизма.

В пользу этого предположения свидетельствуют некоторые литературные данные. В частности, некоторыми авторами (Broadbent, Ladefoged, 1960; Fry et al., 1962; Lindner, 1966) было описано явление контраста, которое заключается в том, что стимул, находящийся вблизи границы между фонемами *A* и *B*, воспринимается как *A*, если он следует за *B*, и как *B*, если следует за *A*. Чтобы проверить, могло ли это явление иметь значение в нашем эксперименте, было проанализировано, каким образом изменяется суждение о стимуле в зависимости от предшествующего стимула. Результаты соответствующей обработки данных, полученных при прослушивании второй таблицы, приведены на рис. 4. Рисунок показывает процент [е]-реакций на разные стимулы при их предъявлении после 3-го, 5-го, 6-го и 8-го стимулов. Можно видеть, что стимулы чаще интерпретируются как гласный [е] в том случае, если предшествующий стимул ближе к гласному [i].

Таким образом, механизм контраста, видимо, имеет место. Однако нельзя с уверенностью сказать, является ли этот механизм единственной причиной обсуждаемых эффектов.

Л и т е р а т у р а

- Г о л у з и н а А. Г. 1971. Шкалирование субъективных расстояний между синтетическими гласными. *Наст. сб.*: 155—160.
- К у з ь м и н Ю. И., Д. М. Л и с е н к о. 1971. Фонетическая интерпретация стимулов с резкими изменениями интенсивности. *Наст. сб.*: 83—92.
- М у ш н и к о в В. Н., Л. А. Ч и с т о в и ч. 1971. О слуховом описании гласного. I. Признаки, различающие [i] и [e]. *Наст. сб.*: 5—10.
- B r o a d b e n t D. E., P. L a d e f o g e d. 1960. Vowel judgements and adaptation level. *Proc. Roy. Soc., B*, 151: 384—399.
- C h i s t o v i c h L. A., G. F a n t, A. d e S e r p a - L e i t a o, P. T j e r n l u n d. 1966. Mimicking of synthetic vowels. *STL—QPSR*, 2: 1—18.
- C r o s s D. V., H. L. L a n e. 1962. On the discriminative control of concurrent responses. The relations among response frequency, latency and topography in auditory generalisation. *J. Exper. Anal. Behav.*, 5: 487—496.
- F r y D. B., A. S. A b r a m s o n, P. D. E i m a s, A. M. L i b e r m a n. 1962. The identification and discrimination of synthetic vowels. *Language a. Speech*, 5: 171—189.

- Fujisaki H., T. Kawashima. 1967. The roles of pitch and higher formants in the perception of vowels. Confer. on Speech Communication a. Processing, Boston : 251—256.
- Ladefoged P., D. E. Broadbent. 1957. Information conveyed by vowels. J. Acoust. Soc. Amer., 29 : 98—104.
- Lindner G. 1966. Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes. Ztschr. f. Phonetik, 19 : 287—307.
- Studdert-Kennedy M., A. M. Liberman, K. N. Stevens. 1963. Reaction time to synthetic stop consonants and vowels at phoneme center and at phoneme boundaries. J. Acoust. Soc. Amer., 35 : 1009.

ОЦЕНКА ГОЛОСА ДИКТОРА ПО ИЗОЛИРОВАННЫМ ГЛАСНЫМ ЗВУКАМ

Р. Л. Шейкин

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

Исследование способности человека различать мужские и женские голоса представляет интерес не только с точки зрения проблемы опознавания дикторов, но и с точки зрения проблемы распознавания фонем. Различия в величине речевого тракта у мужчин и женщин обуславливают то, что частоты формант гласных у женщин оказываются в среднем на 18% более высокими, чем у мужчин (Fant, 1966). Это существенно затрудняет автоматическое распознавание гласных звуков и порождает предположение о том, что человек каким-то образом использует информацию о дикторе для управления положением границ между фонемными категориями. Экспериментальные данные в пользу этой точки зрения были получены в работе Фуджисаки и Кавашима (Fujisaki, Kawashima, 1967).

Настоящая работа основывалась на предположении, что у человека существует четкое представление о типично мужском и типично женском голосе и что он способен для изолированного гласного звука любого диктора определить положение этого гласного на шкале мужской—женский голос. В работе исследовалось, зависит ли такая оценка голоса диктора от фонемного качества гласного и от того, звонкий это гласный или шепотный.

МЕТОДИКА

Для исследования были избраны изолированные гласные звуки [a], [i], [u]. Эти звуки, произнесенные обычным голосом и громким шепотом 22 дикторами-мужчинами и 22 дикторами-женщинами, были записаны на магнитную ленту (магнитофон МЭЗ-15), из этих записей было изготовлено 6 таблиц. Каждая таблица состояла из случайной последовательности одного и того же гласного, произнесенного дикторами-мужчинами и дикторами-женщинами.

Таблицы были предложены для прослушивания 20 аудиторам. Аудиторы должны были решить в отношении каждого звука, насколько он отклоняется от типично мужского или типично женского голоса и выразить свою оценку в баллах по 7-балльной шкале.

Было проведено две серии опытов. В первой из них аудиторам предлагалось оценивать типично мужской голос оценкой в 1 балл, а типично женский голос оценкой в 7 баллов. Во второй серии опытов — наоборот.

РЕЗУЛЬТАТЫ

Начальная обработка материалов опытов показала, что обе серии опытов дают практически одни и те же результаты. Это позволило сделать пересчет результатов, полученных во второй серии опытов, по шкале, принятой в первой серии опытов (через зависимость: $X = 8 - Y$, где X — оценка в первой серии опытов, Y — оценка во второй серии опытов), и объединить материал обеих серий для дальнейшего рассмотрения. При этом оказалось, что разные аудиторы решают задачу как бы двумя различными стилями. В ответах одних аудиторов резко преобладают крайние оценки (1 и 7 баллов). Другие аудиторы при определении близости стимула к типично мужскому или типично женскому голосу применяли весь набор возможных оценок. Преобладания общего числа крайних оценок над общим числом всех других оценок у них не было. Так как это явление приходилось учитывать при обработке материала, то оно было использовано в качестве критерия разделения всех аудиторов по стилю решения задачи на две группы:

	Число крайних оценок	Число остальных оценок	Число правильных оценок
«Неопытные»	2898	2382	4194
«Опытные»	1503	3777	4307

Группы случайно оказались равными по численности. Рассмотрение состава участников этих групп показало, что в первой группе («неопытные») 9 аудиторов из 10 были люди, ранее не участвовавшие в подобных экспериментах. В другой группе аудиторов («опытные») 7 из 10 были сотрудниками лаборатории, часто принимавшими участие в подобных опытах.

Для сравнения обеих групп с точки зрения их способности к распознаванию мужских и женских голосов была проведена следующая обработка данных. Все оценки были разбиты на правильные и ошибочные. Правильными считались оценки 1, 2, 3 балла для мужских и 5, 6, 7 баллов для женских голосов в первой серии опытов, и наоборот — во второй серии опытов. Оценка 4 балла считалась неопределенной; половина общего числа случаев употребления этой оценки была отнесена в категорию правильных, половина — в категорию ошибочных. Число правильных ответов в обеих группах оказалось практически одинаковым. Для проверки того, действительно ли аудиторы из группы «опыт-

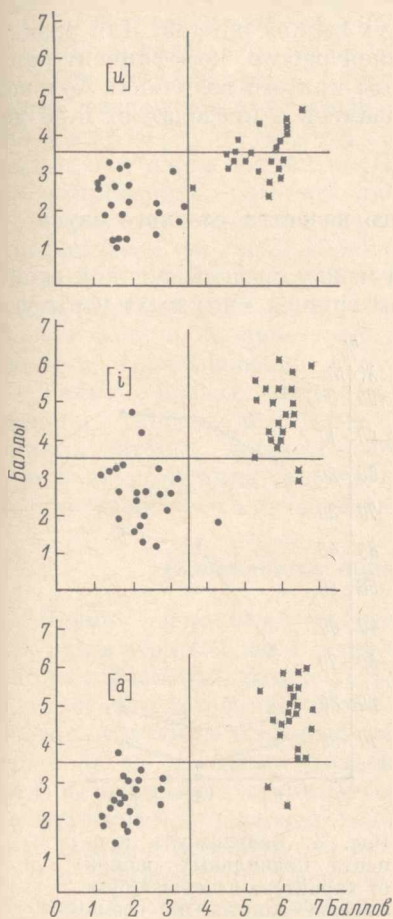


Рис. 1. Соотношение между средней оценкой звонких и шепотных гласных (данные по группе «опытных»).

По оси абсцисс — значение оценки звонкого гласного; по оси ординат — значение оценки шепотного гласного того же диктора. На рис. 1—4: точки — дикторы-мужчины, * — дикторы-женщины, крестики — дикторы-опытные.

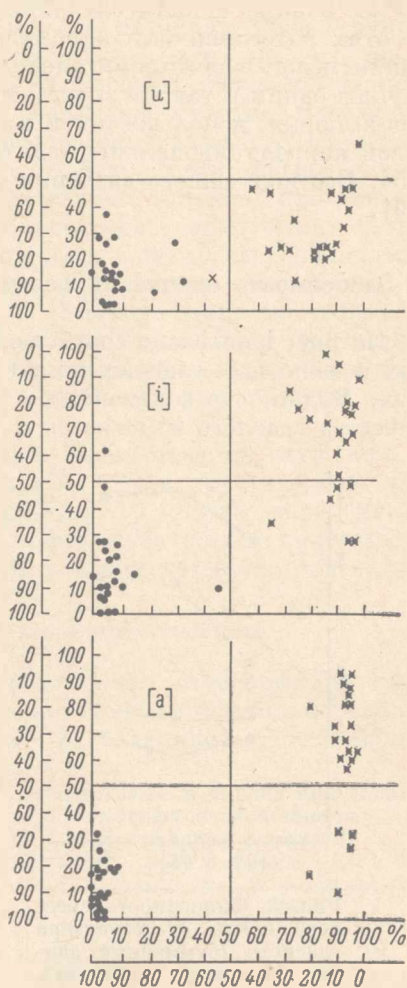


Рис. 2. Характеристика гласных звуков процентом правильных оценок (данные по всему материалу).

По оси абсцисс — процент правильных оценок звонкого гласного; по оси ординат — процент правильных оценок шепотного гласного того же диктора.

ных» использовали какие-то признаки для определения близости стимула к типично мужскому или типично женскому голосу или же их оценки носили случайный характер, в пределах каждой из этих категорий был определен коэффициент ранговой корреляции между значениями оценок, приписанных аудиторами этой группы одним и тем же звукам в двух сериях опытов. Для размеров выборки, использованной в данной работе, коэффициент ранговой корреляции является значимым при его величинах больше 0.54. Все полученные значения оказались в пределах от 0.56 до 0.91.

Зависимость оценки от фонемного качества гласного звука

На рис. 1 показано соотношение между средней оценкой звонких и шепотных гласных по данным группы «опытных» испытуемых. Видно, что с изменением качества гласного изменяется и

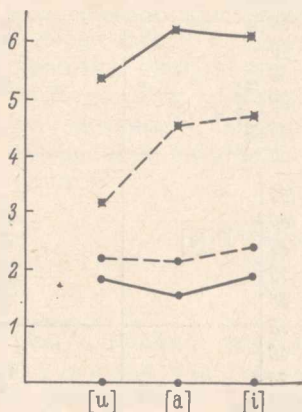


Рис. 3. Зависимость среднего балла от характера гласного (суммарные данные по всему материалу).

По оси абсцисс — фонемы; по оси ординат — средняя оценка, баллы. Сплошная линия — звонкие гласные, прерывистая — шепотные гласные.

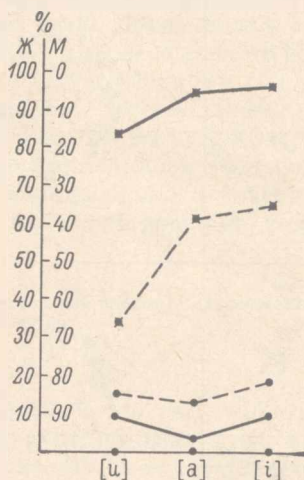


Рис. 4. Зависимость процента правильных оценок от характера гласного (суммарные данные по всему материалу).

По оси абсцисс — фонемы; по оси ординат — процент правильных оценок. Сплошная линия — звонкие гласные, прерывистая — шепотные гласные.

его оценка, причем это наиболее четко выражено для женских шепотных гласных.

На рис. 2 представлена аналогичная, но несколько огрубленная картина (по данным обеих групп), которая наблюдается, если характеризовать гласный звук процентом правильных оце-

нок (способ перехода от баллов к категории правильная—ошибочная оценка описана выше).

На следующих рисунках приведены суммарные данные для дикторов-мужчин и дикторов-женщин (по материалу обеих групп). На рис. 3 показана величина среднего балла, а на рис. 4 — средний процент правильных оценок, полученных каждым гласным.

Можно видеть, что наилучшее различение мужских и женских голосов наблюдается в случае произнесения гласного [a]. Хуже всего распознается гласный [u], особенно если он шепотный. Из рис. 1 и 2 следует, что оценки разных гласных, произнесенных одним и тем же диктором, не являются одинаковыми. Следовало проверить, остается ли при этом неизменным положение дикторов относительно друг друга. Для этого был вычислен коэффициент ранговой корреляции r^s между оценками, полученными для разных гласных одного и того же диктора. Как оказалось, только в двух случаях из 18 значения r^s достигли величины 0.56 и 0.64, в то время как для данной выборки значимы величины r^s , превышающие 0.54.

Таким образом, аудиторы при распознавании мужских и женских голосов используют, видимо, такие акустические признаки, которые являются функциями от многих переменных, а не только от таких связанных с полом особенностей, как механические свойства голосовых связок и длина речевого тракта.

Особенности оценки шепотных гласных

Оценка шепотных гласных представляет особый интерес, так как в этом случае в сигнале может сохраниться только информация о размерах речевого тракта и не сохраняется информация о голосовом источнике (особенностях гортани). Казалось бы, что переход на шепот должен был бы одинаково сказываться на мужских и женских голосах. Однако, как видно из рис. 1, 2 и таблицы, число женских шепотных гласных, воспринимаемых как мужские, существенно превышает число мужских гласных, воспринимаемых как женские.

Этот факт пока трудно объяснить, однако можно высказать предположение, что с точки зрения каких-то измерений, производимых слухом, шумовое возбуждение, используемое при шепоте, более близко к голосовому возбуждению мужских голосов, чем к голосовому возбуждению женских голосов. Физическим при-

Восприятие и оценка шепотных гласных в зависимости от качества фонемы и пола диктора

Передано	Принято					
	[u]		[a]		[i]	
	М	Ж	М	Ж	М	Ж
М . . .	44	0	44	0	41	3
Ж . . .	37	7	11	38	9	35

знаком, дающим такую информацию, могло бы быть расстояние между составляющими в спектре сигнала. В случае шепотного гласного воспринимаемый слухом шумовой сигнал имеет сплошной спектр. В спектре мужского голоса частотная разница между гармониками мала. В спектре женских голосов это расстояние велико — сигнал приближается к тональному. То обстоятельство, что при произношении шепотом мужские и женские гласные

как-то различаются, позволяет считать, что форма спектра (частоты формант) является информацией, используемой слухом для оценки половой принадлежности диктора.

Рассмотрим два предположения о связи частот формант с этой оценкой. Одно из них состоит в том, что оценка основывается на абсолютных значениях частот формант. В таком случае это может касаться только третьей и более высоких формант. Однако данные Фанта о частотах формант мужских и женских голосов свидетельствуют, что по абсолютным значениям частот формант определить принадлежность фонем к женскому или мужскому голосам не удастся, так как эти значения перекрываются. В частности, это явствует из приведенных ниже значений частоты (Гц) третьей форманты:

	[u]	[a]	[i]
М . . .	2230	2520	2960
Ж . . .	—	2820	3433

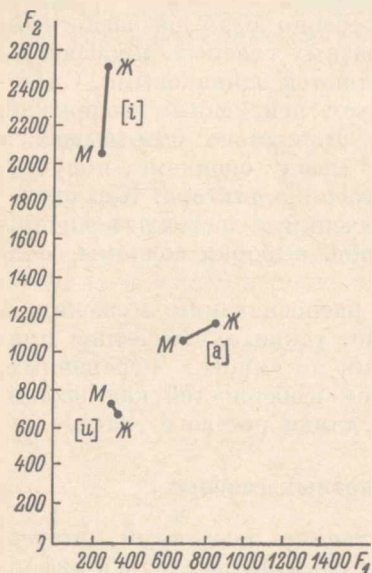


Рис. 5. Векторные расстояния между мужскими и женскими голосами (Fant, 1966).

По оси абсцисс — частота первой форманты; по оси ординат — частота второй форманты.

Тогда остается другое предположение, а именно то, что человек оценивает не абсолютные значения частот формант, безотносительно к воспринимаемой фонеме, а положение частот формант внутри фонемной категории (области). В таком случае можно допустить, что основная информация заключена в частотах первой и второй формант. Как показал Слоусон (Slawson, 1968), эти форманты играют определенную роль в оценке тембра. На рис. 5 приведены построенные по данным Фанта значения частот F_1 и F_2 гласных [u], [a], [i] для мужских и женских голосов. Видно, что векторное расстояние между мужскими и женскими голосами для гласного [u] значительно меньше, чем для

других гласных, и что по F_2 мужское [u] может быть даже выше женского [u]. Это может в какой-то мере объяснить трудность в опознавании шепотного гласного [u], имевшую место в данной работе.

ВЫВОДЫ

1. Отнесение изолированного гласного звука к мужскому или женскому голосам зависит от фонемного качества гласного. Из трех исследованных гласных [u], [a], [i] наименее благоприятным для распознавания голоса является гласный [u]. Зависимость оценки голоса диктора от фонемного качества более отчетливо выражена в случае шепотных гласных.

2. Характерной ошибкой распознавания является то, что женские голоса принимаются за мужские. Обратные ошибки наблюдаются относительно редко.

Литература

- Fant G. 1966. A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transm. Lab. Quart. Progr. a. Status Report*, 4 : 22—30.
- Fujisaki H., T. Kawashima. 1967. The roles of pitch and higher formants in the perception of vowels. *Confer. of Speech Communication a. Processing*, Boston : 251—256.
- Slawson A. W. 1968. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *J. Acoust. Soc. Amer.*, 43 : 87—101.

ШКАЛИРОВАНИЕ СУБЪЕКТИВНЫХ РАССТОЯНИЙ МЕЖДУ СИНТЕТИЧЕСКИМИ ГЛАСНЫМИ

А. Г. Голузина

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

Задачей настоящей работы было выяснение того, каким образом человек производит сравнение синтетических гласных при определении степени различия между ними. Можно высказать две гипотезы.

1. Сравнение звуков происходит в пространстве слуховых образов. Степень различия между двумя звуками определяется как расстояние между соответствующими им точками в этом пространстве.

2. Сравнение звуков происходит на более поздней ступени восприятия, когда звукам приписаны символы.

Рассмотрим, какие результаты должны получаться в экспериментах по шкалированию субъективных расстояний между син-

тетическими гласными в том и другом случае. Для этого удобно представить восприятие гласного в виде следующей модели.

Пусть испытуемому предъявляется гласный S , который характеризуется в r -мерном пространстве физических признаков точкой с координатами $F_1, F_2, \dots, F_r$, где F_i — i -я формантная частота звука. Слуховой образ звука S представляется точкой A в пространстве слуховых образов. Из-за влияния случайных эффектов множество слуховых образов данного звука будет образовывать некоторое распределение со средним $\bar{A}$. Пространство слуховых образов разделено на множество m областей $D_1, D_2, \dots, D_m$. Если слуховой образ звука S попадает в область D_j , звуку приписывается символ C_j .

Пусть $S_1, S_2, \dots, S_n$ — набор гласных, распределения слуховых образов которых последовательно располагаются в слуховом пространстве вдоль некоторой траектории, проходящей через области $D_{j_1}, D_{j_2}, \dots, D_{j_k}$ (области $D_{j_h}, h=1, 2, \dots, k$, из множества m и траектория выбраны так, что расстояния между средними $\bar{A}_1$ и $\bar{A}_i$ распределений, соответствующих звукам S_1 и S_i , растут с увеличением i), причем располагаются достаточно часто. Будем предъявлять испытуемому пары гласных $S_1 S_i$ ($i=1, 2, \dots, n$) для оценки степени различия между ними.

Отвлекаясь пока от случайных флюктуаций слуховых образов, рассмотрим, как будут изменяться субъективные расстояния между гласными для последовательных пар $S_1 S_i$ ($i=1, 2, \dots, n$), если при их оценке испытуемый ведет себя в соответствии с одной из гипотез. Согласно первой гипотезе, расстояния будут измеряться в пространстве слуховых образов и, следовательно, будут большими при больших i . По второй гипотезе субъективные расстояния при увеличении i будут оставаться неизменными до тех пор, пока слуховые образы A_i гласных S_i попадают в одну и ту же область D_{j_h} , т. е. S_i отождествляются с одним и тем же символом C_{j_h} , и возрастут скачком, как только A_i попадет в соседнюю область $D_{j_{h+1}}$. Таким образом, субъективные расстояния с увеличением i будут возрастать ступенчато.

Теперь, учитывая случайные флюктуации слуховых образов гласных, будем рассматривать среднее расстояние для пары гласных. Если справедлива первая гипотеза, субъективные расстояния по-прежнему будут расти с ростом i . Если же верна вторая, ступенчатый характер изменения субъективного расстояния будет нарушаться в большей или меньшей степени в зависимости от величины дисперсии слуховых образов.

Таким, образом, если флюктуации слуховых образов гласных достаточно малы, по характеру изменения субъективных расстояний между звуками в последовательных парах $S_1 S_i$ ($i=1, 2, \dots, n$) можно установить, какая гипотеза верна. Если верна вторая

гипотеза, следует ожидать также, что дисперсия субъективного расстояния между S_1 и S_i будет заметно возрастать для тех пар, у которых S_i находится вблизи категорной границы.

МЕТОДИКА

Чтобы определить, какая из двух гипотез верна, был проведен следующий эксперимент. Были использованы 12 синтетических гласных, расположение которых на плоскости (F_1 , F_2) показано на рис. 1. Подробная характеристика этих гласных дана в работе Чистович с соавторами (Chistovich et al., 1966). По звучанию эта последовательность сигналов представляет собой постепенный переход гласного [i] в гласный [a]. Различные пары этих сигналов (всего 144 пары) были записаны в случайном порядке на магнитную ленту. Длительность каждого сигнала была 300 мсек. Запись предъявляли испытуемым и предлагали им оценить степень различия между сигналами в паре в баллах от 0 до 7, приписывая 0 паре одинаковых сигналов, а 7 — паре [a]—[i]. В эксперименте принимали участие 30 испытуемых, которые прослушивали ленту от одного до трех раз (повторные прослушивания проходили с интервалом от нескольких часов до нескольких суток). Всего было получено по 56 оценок для каждой пары.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

На основании ответов испытуемых для каждой предъявляемой пары сигналов было получено частотное распределение оценок расстояния между сигналами в баллах от 0 до 7. Затем было произведено сравнение распределения для пары различных сигналов $S_i S_j$ с распределением для этой же пары сигналов, предъявляемых в обратном порядке: $S_j S_i$. Для сравнения распределений применялся критерий χ^2 с уровнем значимости 5%. Для 12 из 64 пар гипотеза однородности распределений была отвергнута. Следовательно, распределения оценок расстояний для пар $S_i S_j$ и $S_j S_i$ нельзя считать однородными, и они не могут быть объединены. Далее для каждой из 144 пар сигналов было вычислено среднее и дисперсия полученных оценок расстояния между сигналами. Средние и дисперсии для последовательных пар $S_i S_j$ ($j=1, 2, \dots, 12$) при некоторых i приведены соответственно на рис. 2 и 3.

Из рис. 2 видно, что для последовательных пар $S_1 S_j$, $S_2 S_j$, $S_{11} S_j$, $S_{12} S_j$ ($j=1, 2, \dots, 12$) расстояние изменяется медленно

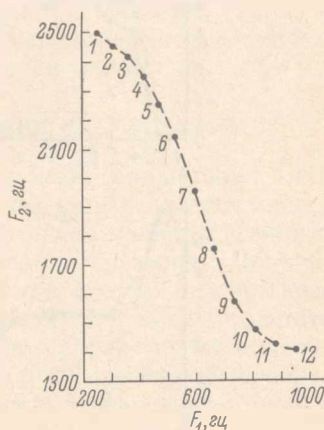


Рис. 1. Расположение синтетических гласных, использованных в эксперименте, на плоскости $F_1 F_2$.

По оси абсцисс — частота первой форманты F_1 , Гц; по оси ординат — частота второй форманты F_2 , Гц.

на участке от $j=4$ до $j=9$ и быстро на участках от $j=2$ до $j=4$ и от $j=9$ до $j=11$. Дисперсии имеют пики при $j=3$, $j=4$ и $j=9$, $j=10$ (рис. 3). Из этого можно заключить, что измерение субъективного расстояния между гласными происходит после того, как произведена классификация звуков, т. е. подтверждается вторая гипотеза. Звуки S_1, S_2, S_3 можно отнести к одному классу,

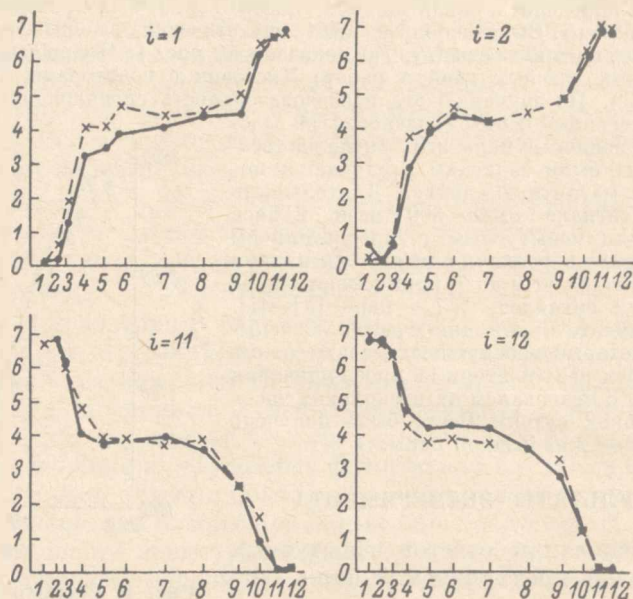


Рис. 2. Средние значения оценок расстояний между гласными S_i ($i=1, 2, 11, 12$) и каждым из остальных гласных.

Сплошные линии — данные для пар $S_i S_j$, прерывистые — для пар $S_j S_i$ ($j=1, 2, \dots, 12$). По оси абсцисс — номера j , расположенные соответственно расстоянию S_j от S_i на двухформантной плоскости; по оси ординат — средняя оценка.

$S_4, S_5, S_6, S_7, S_8, S_9$ — ко второму классу, S_{10}, S_{11}, S_{12} — к третьему классу. Чтобы определить, что собой представляют эти классы, на группе испытуемых из 9 человек был проведен эксперимент по записи буквами русского языка прослушиваемых синтетических гласных $S_1, S_2, \dots, S_{12}$, предъявляемых в случайном порядке. Результаты эксперимента, приведенные на рис. 4, показывают, что сигналы делятся на три класса: S_1-S_3, S_4-S_8 и S_9-S_{12} . Сравнивая эти классы с классами, полученными при шкалировании субъективных расстояний, можно заключить, что последние соответствуют фонемам национального языка. Следовательно, при шкалировании субъективных расстояний происходит оценка расстояний между фонемами. На рис. 5 приведены сред-

ние расстояния в баллах для тех пар $S_i S_j$, в которых оба сигнала принадлежат ко второму классу. Эти расстояния изменяются

значительно. Это показывает, что если сигналы в паре относятся к одному классу, то измерение расстояния происходит другим способом: используется более подробное описание сигналов.

Из результатов эксперимента можно сделать вывод, что сравнение гласных происходит следующим образом: если гласные принадлежат к различным классам, то сравниваются классы, если к одному классу, то сравниваются более подробные описания сигналов.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные данные показывают, что при определении степени различия между гласными человек запоминает и фонему, и подробное описание сигналов. Подробное описание используется для сравнения только в том случае, когда гласные принадлежат к одному фонемному классу. Такой же характер восприятия был обнаружен Фуджисаки и Кавашимой в экспе-

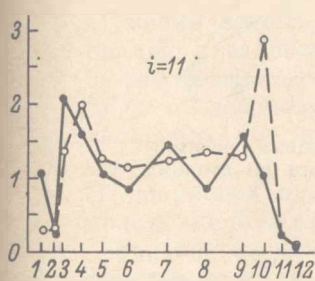
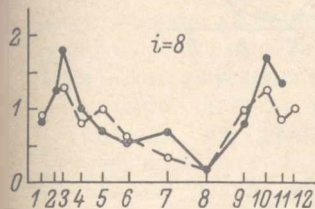
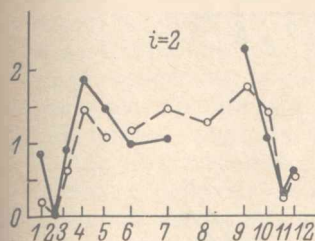
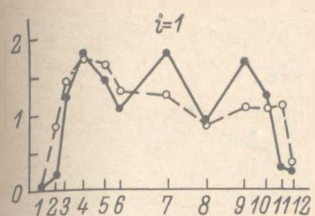


Рис. 3. Дисперсии оценок субъективных расстояний между гласными в предъявляемых парах (по оси ординат).

Обозначения те же, что на рис. 2.

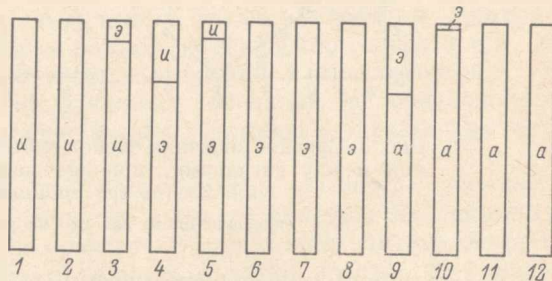


Рис. 4. Процентное соотношение разных букв, использованных при записи синтетических гласных буквами русского алфавита.

Арабские цифры — номера стимулов в соответствии с рис. 1.

риментах по АВХ-различению (Fujisaki, Kawashima, 1968).

Результаты проведенного исследования позволяют заключить, что измерение субъективного расстояния происходит после того,

как произведена фонемная классификация гласных. Тогда применение метода многомерного шкалирования к гласным данного языка (если они принадлежат к различным фонемным классам)

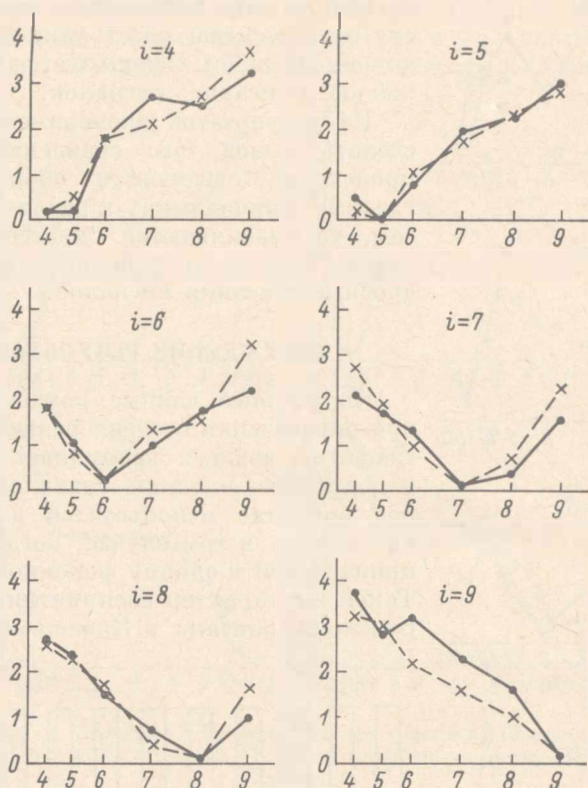


Рис. 5. Средние значения оценок субъективных расстояний между сигналами, принадлежащими ко второму классу (по оси ординат).

Обозначения те же, что на рис. 2.

не может помочь при изучении восприятия этих гласных, так как измерение субъективных расстояний происходит после того, как процесс восприятия окончился.

Литература

- Chistovich L. A., G. Fant, A. de Serpa-Leitao, P. Tjernlund. 1966. Mimicking of synthetic vowels. STL-QPSR, 2: 1—18.
 Fujisaki H., T. Kawashima. 1968. The influence of various factors on the identification and discrimination of synthetic speech sounds. Reports of 6th Intern. Congr. on Acoustics, Tokyo, II, B: 95—97.

ОБ ИНТОНАЦИОННОМ ЧЛЕНЕНИИ НЕКОТОРЫХ ТИПОВ ПРЕДЛОЖЕНИЙ РУССКОГО ЯЗЫКА

О. Ф. Кривнова, А. В. Венцов

Лаборатория прикладной и структурной лингвистики МГУ;
Институт физиологии им. И. П. Павлова АН СССР, Ленинград

В последнее время изучению интонации уделяется значительное внимание как исследователями синтаксиса, так и фонетистами (Chomsky et al., 1956; Stockwell, 1960; Bierwish, 1966; Isačenko, Shadlich, 1966; Ohman, 1967). Если первые на основе синтаксических и других особенностей языка пытаются создать системы правил для управления интонацией и с их помощью объяснить фонетические особенности произнесения различных типов предложений, то вторые, изучая фонетические особенности речи, стремятся выяснить, какими явлениями более высоких уровней языка они вызываются. Характерно при этом, что и те, и другие рассматривают интонацию не просто как некоторое акустическое явление, а как абстрактную структуру, являющуюся своего рода программой порождения акустического сигнала. При этом акустическим коррелятом интонации обычно считают изменение частоты основного тона $F_0(t)$.

Исходя из детального анализа интонационных контуров, многие авторы высказывают предположение, что интонация как некоторое результирующее явление определяется следующими факторами:

- 1) общим интонационным типом, устанавливающим основную интонацию предложения — вопрос, утверждение, эмпфаза и т. д.;
- 2) распределением словесных ударений в предложении;
- 3) распределением пограничных сигналов в предложении, разделяющих его на более мелкие единицы — фразы или синтагмы.

В настоящее время особенно большое внимание уделяется изучению влияния на интонацию третьего фактора. По-видимому, это объясняется тем, что вопрос о принципах интонационного членения может оказаться важным с точки зрения, во-первых, исследования связи фонетического и более высоких уровней языка и, во-вторых, для решения общей проблемы членения речевого сигнала при автоматическом распознавании речи.

В лингвистических исследованиях высказываются различные мнения о причинах, определяющих интонационное членение предложения. Одни авторы считают, что разграничительная функция интонации обусловлена синтаксической структурой предложения, а интонационно выделяемые группы являются его синтаксическими составляющими (Wells, 1947; Trager, Smith, 1951; Lieberman, 1967). Другие отрицают подобное влияние синтакси-

ческой структуры предложения на интонацию и предполагают, что разграничительная функция интонации подчинена принципу, в соответствии с которым интонационно выделяемые группы едины прежде всего с точки зрения коммуникации и не обязательно должны совпадать с синтаксическими составляющими предложения. При таком подходе объединение слов в одну интонационную единицу обусловлено такими факторами, как контекстные условия, намерение говорящего и т. д. (Karcevskij, 1964). Такое расхождение в точках зрения, по-видимому, вызвано недостаточно четким выбором критерия при членении предложений на интонационные единицы. Чаще всего для членения используется критерий фонетической оформленности интонационной единицы.

В русской фонетике выделяемые в предложении интонационные единицы обычно называются синтагмами (Щерба, 1937; Зиндер, 1960). Считается, что синтагма является интонационно неделимой единицей речевого потока, а главной ее характеристикой служит синтагматическое ударение. Основным средством выражения такого синтагматического ударения, заключающегося в усилении степени ударенности последнего слова синтагмы, считается мелодика, т. е. повышение или понижение высоты голоса на этом ударном слове (Щерба, 1937; Златоустова, 1962).

Высказывается мнение о существовании двух важнейших типов синтагматического ударения: для первого характерно низкое движение тона до гласного, на котором реализуется ударение, и резкий подъем тона на этом гласном; для второго — высокое движение тона до гласного и резкий спад на самом ударном гласном (Брызгунова, 1963; Панов, 1968). Отсюда делается вывод, что синтагма выделяется противопоставлением уровней частоты основного тона голоса на пред- и ударной частях синтагмы. Следует, однако, отметить, что экспериментальные данные, на которых основывается такой вывод, нельзя признать достаточно надежными из-за разнородности рассмотренного языкового материала.

В связи с этим мы ставили своей задачей подробно исследовать фонетическую характеристику интонации и интонационное членение некоторых типов предложений русского языка с максимальным учетом специфики интонации как языкового явления.

В качестве исходных были исследованы предложения с синтаксической структурой подлежащее—сказуемое—дополнение (П—С—Д), в которых подлежащее выражено существительным (личным местоимением) в именительном падеже, сказуемое — личным глаголом в изъявительном наклонении и дополнение — существительным (личным местоимением) в винительном падеже. Например, предложение *Немой молил меня*.

Предложения именно этого типа были выбраны из следующих соображений. Во-первых, они являются наиболее распростра-

ненным типом простых предложений в русском языке; во-вторых, относительно их синтагматического членения нет единого мнения. Считается, что чаще всего они реализуются в виде одной синтагмы (Брызгунова, 1963; Златоустова, 1962), но в некоторых случаях возможно членение таких предложений на две синтагмы: П и С—Д (Златоустова, 1962).

МЕТОДИКА

Поскольку акустическим коррелятом интонации обычно считают изменения частоты основного тона голоса, для исследования интонации необходимо непрерывно регистрировать частоту или период колебаний голосовых связок.

В описываемых ниже исследованиях колебания голосовых связок регистрировались с помощью гортанного микрофона, помещенного в яремную впадину испытуемого. Затем с помощью специального прибора, интонометра, длительность периода колебаний связок преобразовывалась в амплитуду напряжения пилообразной формы, которое записывалось на фотопленку с помощью магнитоэлектрического осциллографа Н-102. Огибающая этого сигнала представляет собой кривую зависимости периода основного тона от времени $T_0(t)$.

Сопоставление характерных особенностей изменения кривой $T_0(t)$ с артикуляторными параметрами, а также членение произносимых предложений на слова и слоги обеспечивалось одновременной регистрацией таких явлений, как смыкание и размыкание губ, смыкание языка с твердым нёбом, движения нёбной занавески и др., с использованием специальных датчиков и аппаратуры (Чистович и др., 1965).

Речевой материал был подобран таким образом, чтобы в максимальной степени исключить влияние артикуляции на изменения частоты основного тона: использовались слова, содержащие согласные звуки, произнесение которых практически не сопровождается увеличением внутриротового давления. В тех случаях, когда не удавалось подобрать нужный речевой материал, одновременно с записью изменений периода колебаний голосовых связок регистрировалось внутриротовое давление и вводились соответствующие поправки в кривую $T_0(t)$ (Кривнова, 1969).

Полученные таким образом кривые изменения периода основного тона во времени нормировались по времени и громкости, и на основании нормированных кривых вычислялась зависимость $\lg F_0(t)$, именуемая в дальнейшем интонационной кривой высказывания (Кривнова, 1969).

РЕЗУЛЬТАТЫ

В предварительных экспериментах были проанализированы интонационные кривые 30 предложений, произнесенных тремя испытуемыми с повествовательной интонацией (всего около 500 произнесений). Оказалось, что во всех без исключения произнесениях интонационная кривая предложения может быть схематично представлена в виде кривой, приведенной на рис. 1, которая характеризуется следующими особенностями:

1) четко выделяются два подъема частоты основного тона (интервалы T_1T_2 и T_5T_6) и два спада (интервалы T_3T_4 и T_7T_8);

2) за каждым подъемом частоты следует спад, так что следующий подъем начинается примерно от того же уровня частоты основного тона, что и предыдущий;

3) в интервале T_6T_7 понижения частоты основного тона не наблюдается.

Возникает естественный вопрос, какова программа этих изменений и с реализацией каких элементов предложения она связана.

Сразу же возникает предположение, что наблюдающиеся изменения частоты основного тона служат для интонационного выражения словесного ударения, так как большинство изменений в интонационной кривой реализуется именно на ударных слогах слов, входящих в предложение (рис. 1).

Однако такая точка зрения встречает несколько серьезных возражений. Если указанные изменения частоты основного тона служат для выражения только словесного ударения, следует

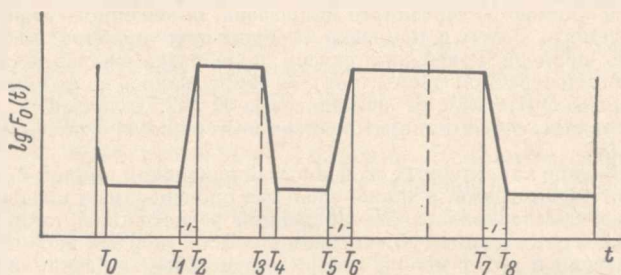


Рис. 1. Схематическое изображение интонационного контура предложения типа подлежащее—сказуемое—дополнение.

По оси ординат — логарифм частоты основного тона. Сплошные вертикальные линии — границы предложения, прерывистые — границы слов; квадратные скобки — ударные слоги.

ожидать, что во всех словах они окажутся одинаковыми и будут реализоваться в пределах ударного слога, так как заданы в его программе. Но в последнем слове предложения, в отличие от первых двух слов, на ударном слоге происходит не подъем, а падение частоты основного тона. Учитывая известный факт интонационного оформления абсолютного конца высказывания понижением частоты основного тона, логично предположить, что именно эту функцию выполняет снижение частоты на ударном слоге последнего слова в исследуемых предложениях. В этих условиях изменение частоты основного тона на этом ударном слоге должно задаваться программой всего предложения, а не ударного слога.

Далее. Подъем частоты на ударном слоге первого слова предложения сопровождается ее падением примерно до того уровня, с которого начался подъем. Если бы эти изменения частоты служили только для выражения словесного ударения и задавались программой реализации ударного слога, следовало бы ожидать, что длительность стационарного участка T_2T_3 (рис. 1) не будет

превышать длительности ударного слога. Однако на всех экспериментально полученных интонационных кривых указанное падение частоты начинается, как правило, с началом второго слова. При этом длительность интервала T_2T_3 возрастает с увеличением числа слогов в заударной части первого слова (рис. 2). Следовательно, эти изменения частоты не могут задаваться программой одного ударного слога, а определяются программой более крупного элемента (в данном случае всего первого слова).

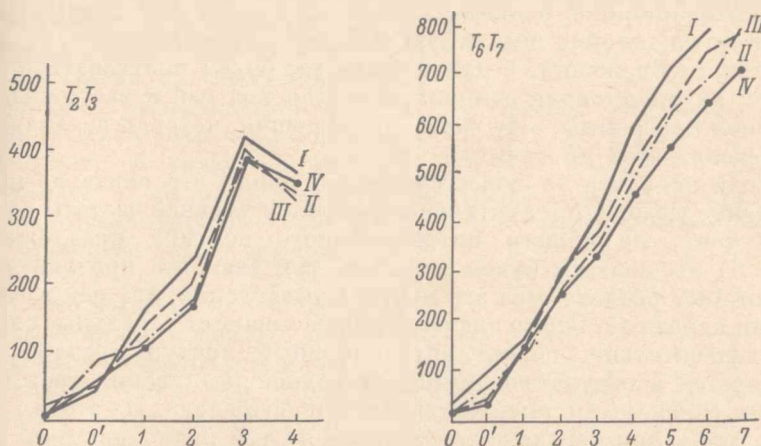


Рис. 2. Зависимость длительности стационарных участков (T_2T_3 , T_6T_7) интонационной кривой (по оси ординат, мсек.) от числа слогов в соответствующей части слова (по оси абсцисс).

Нуль соответствует случаю, когда исследуемый интервал содержит только ударный гласный, O' — исследуемый интервал содержит ударный гласный с согласным перед ним или после него. Римскими цифрами обозначены результаты для разных дикторов.

Наконец, подъем частоты основного тона на ударном слоге второго слова в предложении не сопровождается падением частоты ни в пределах ударного слога, ни в пределах самого слова. Снижение частоты происходит только на ударном слоге конечного слова предложения, при этом длительность стационарной части участка интонационной кривой между ударными слогами этих слов (интервал T_6T_7) является функцией суммы числа слогов в заударной части второго слова и предударной последнего (рис. 2). Очевидно, этот участок интонационной кривой образуется при реализации программы двух последних слов.

Из сказанного выше мы можем сделать вывод, что изменения частоты основного тона при произнесении предложения типа П—С—Д служат для интонационного выражения некоторых элементов предложения, а не словесного ударения входящих в него

слов, хотя и реализуются при произнесении именно ударных слов. Попытаемся выяснить, каковы эти элементы и из каких частей складывается программа реализации интонационной кривой такого предложения.

Если интонационный контур целого предложения может быть получен в результате реализации нескольких интонационных программ, привязанных к определенным членам предложения, естественно ожидать, что часть предложения, произносимая одновременно с реализацией одной из этих программ, окажется синтагмой, или интонационной единицей, поскольку последняя представляет собой интонационно неделимую единицу речевого потока. Исходя из этих соображений, в дальнейшем мы будем пользоваться термином интонационная единица, понимая под ней и часть интонационной программы, и ту часть предложения, которая реализуется в пределах этой программы.

Если исходить из существующего мнения, что синтагма интонационно выделяется противопоставлением уровней частоты основного тона, из общего интонационного контура предложения (рис. 1) мы должны будем выделить три элемента предложения, на которых реализуются эти противопоставления: два неконечных и один конечный. Этими элементами оказываются отдельные слова, и, следовательно, общий интонационный контур предложения образуется в результате последовательной реализации трех программ, связанных с произнесением каждого слова. Из этого вытекает, что такие предложения делятся на три интонационные единицы (рис. 3, а). Однако такое членение сразу встречает серьезные возражения: во-первых, в литературе нет никаких сведений о возможности членения такого типа предложений на три синтагмы; во-вторых, невозможно установить границу между конечной и предшествующей ей неконечной интонационными единицами, не прибегая к анализу семантико-синтаксических связей между словами в соответствующем отрезке предложения.

Вместе с тем исследуемые интонационные кривые имеют одну характерную особенность, которая дает возможность судить о способе задания наблюдаемых изменений частоты основного тона: за каждым подъемом частоты основного тона следует ее падение, причем примерно до того уровня, с какого начался подъем. Это позволяет предположить, что элемент программы, определяющей полный интонационный контур предложения, состоит из одного импульса изменения частоты, а интонационный контур предложения типа П—С—Д образуется реализацией двух элементов программы и состоит из двух единиц: одной неконечной и одной конечной (рис. 3, б). Конечную единицу образуют два последних слова в предложении, в данном случае — сказуемое и дополнение, о составе неконечной единицы можно только сказать, что в данном случае она состоит из одного подлежащего. В этом случае предложения такого типа разделяются на две интонационные

единицы: П и С—Д, что не противоречит существующим в литературе мнениям о членении этого типа предложений на синтагмы, а границу между единицами легко установить по полной интонационной кривой предложения.

Однако несмотря на кажущуюся предпочтительность этого способа задания интонационной программы и членения предложения на синтагмы, нам могут возразить, что предлагаемая конечная единица на самом деле состоит из двух самостоятельных единиц: одной неконечной и одной конечной, что сказуемое и дополнение все же образуют две отдельные синтагмы и что синтагма интонационно выражается противопоставлением уровней частоты основного тона. В связи с этим возникает необходимость экспериментальной проверки правильности наших предположений.

Ниже мы попытаемся экспериментально показать, что элементом интонационной программы является один импульс изменения частоты основного тона и что такому элементу программы соответствует членение предложения на интонационные единицы, при котором в конечную единицу входят два последних слова в предложении.

Для этого исходное предложение типа П—С—Д было распространено введением в него такого нового члена, который не может образовать самостоятельной интонационной единицы. Существует мнение, что таким членом предложения может быть определение (О), которое из-за своей семантико-синтаксической несамостоятельности будет входить в одну интонационную единицу с подлежащим или дополнением (Златоустова, 1962). Тогда в предложении типа П—С—О—Д, где определение выражено прилагательным, в конечную интонационную единицу должны войти определение и дополнение, а интонационные контуры этой единицы, соответствующие двум описанным выше способам членения предложения на интонационные единицы, будут иметь вид, приведенный на рис. 4. Совершенно очевидно, что по виду интонационной кривой, полученной экспериментально для этой группы,

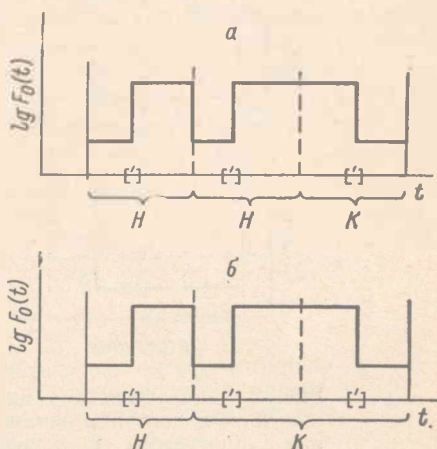


Рис. 3. Два способа членения интонационной программы на более мелкие элементы и вытекающие из них способы членения предложения на интонационные единицы.

а — интонационная единица выделяется противопоставлением уровней частоты основного тона; б — импульсным изменением частоты. Остальные обозначения те же, что на рис. 1.

мы сможем судить о том, каков на самом деле элемент интонационной программы.

Следует, однако, иметь в виду, что экспериментальных данных, подтверждающих семантико-синтаксическую несамостоятельность определения, в настоящее время не существует. Поэтому целесообразно начать с рассмотрения интонационных рисунков некоей единицы, содержащей определение. Если предположение о несамостоятельности определения справедливо, введение его

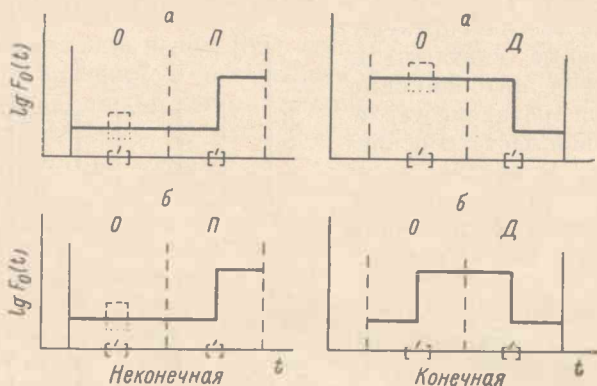


Рис. 4. Гипотетические интонационные контуры не-конечной и конечной единиц при двух способах задания программы и членения предложений типа О—П—С—Д и П—С—О—Д на интонационные единицы.

а — при задании интонации противопоставлением уровней частоты; б — при задании программой импульса изменения частоты. Прерывистыми и точечными линиями отмечены возможные контуры изменения частоты основного тона, связанного с реализацией словесного ударения. Остальные обозначения те же, что на рис. 1.

в интонационную единицу не должно изменить ее интонационного рисунка, который будет одинаковым при обоих способах членения (рис. 4).

Для проверки высказанных здесь предположений были изучены интонационные рисунки 18 предложений типа О—П—С—Д и П—С—О—Д, произнесенных тремя дикторами тоном обычного сообщения (например: *Милая мама мыла Маню, Мама мыла маленькую Маню*). Всего было проанализировано около 70 произнесений. Оказалось, что изменения частоты основного тона в интересующих нас единицах для всех произнесений могут быть схематично представлены следующим образом (рис. 5).

Рассмотрим сначала интонационный контур группы определение—подлежащее (рис. 5, а). Судя по полученному интонационному контуру, эта группа действительно образует одну интонационную единицу. Сомнение может вызвать лишь наблюдающийся

во всех произнесениях подъем частоты основного тона на ударном слоге определения. Однако весьма вероятно, что этот подъем частоты служит для выражения словесного ударения и не относится к интонационной характеристике всей группы. В этом случае наблюдаемое изменение частоты должно реализоваться в пределах ударного слога, и длительности интервалов T_2T_3 и T_3T_4 должны определяться длительностью ударного слога и заударной части определения соответственно, т. е. длительности этих участков интонационной кривой должны возрастать при увеличении длительности соответствующих участков определения. Так как в нормальной речи длительности этих интервалов подвер-

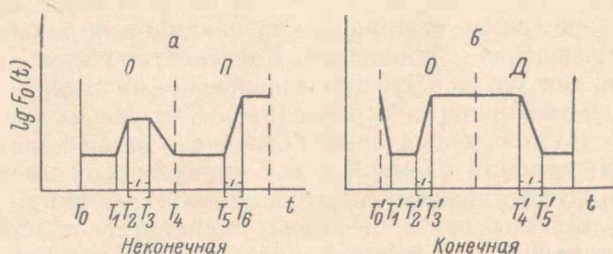


Рис. 5. Схематическое изображение экспериментально полученных интонационных контуров групп определение—подлежащее (а) и определение—дополнение (б).

Обозначения те же, что на рис. 1.

жены естественным флюктуациям, наличие связи между этими величинами и ее степень можно установить, вычислив соответствующие коэффициенты корреляции. Ниже приводим коэффициенты корреляции группы О—П:

	$T_2T_3/T_0^{уг}$	T_3T_4/T_0^3
Диктор I . . .	0.44	0.98
Диктор II . . .	0.38	0.95
Диктор III . . .	0.79	0.97

Здесь $T_0^{уг}$ — длительность ударного гласного в определении и T_0^3 — длительность заударной части определения.

Полученные коэффициенты корреляции говорят о том, что длительность интервала T_3T_4 является линейной функцией длительности заударной части определения, а длительность стационарной части T_2T_3 в основном определяется длительностью ударного гласного определения. В целом на основании этих данных можно с уверенностью утверждать, что наблюдаемые изменения частоты основного тона при произнесении определения в группе

определение—подлежащее представляют собой локальное явление и служат для интонационного выделения словесного ударения. В пользу такого вывода говорит и меньшая амплитуда этих изменений частоты. Приращение частоты основного тона при интонационном выделении словесного ударения составляло в наших экспериментах от 20—30 до 40—50 гц у разных испытуемых, тогда как при интонационном выделении целой единицы оно составляло у тех же испытуемых от 90—110 до 100—115 гц.

Таким образом, мы установили, что в группе О—П определение действительно не образует самостоятельной интонационной единицы, а произнесение его ударного слога сопровождается локализованным в пределах этого слога изменением частоты основного тона.

Обратимся теперь к анализу интонационных характеристик группы определение—дополнение. Как следует из рис. 5, б, интонационный контур этой группы совершенно аналогичен контуру конечной интонационной единицы (рис. 3, б). Необходимо только убедиться, что понижение частоты основного тона во время произнесения определения не служит для выражения словесного ударения, а выделяет конечную интонационную единицу. Если это изменение частоты основного тона является составной частью общего интонационного контура конечной единицы, длительность интервала $T'_1T'_2$ должна увеличиваться при увеличении длительности предударной части определения. В противном случае при интонационном оформлении словесного ударения с помощью понижения частоты основного тона длительность этого интервала не будет превышать длительность ударного слога.

Анализ экспериментально полученных интонационных контуров группы определение—дополнение показал, что длительность интервала $T'_1T'_2$ в интонационном контуре этой группы возрастает при увеличении числа слогов в предударной части определения, а значит и длительности всей предударной части. При этом коэффициенты корреляции между соответствующими длительностями достаточно велики и составляют 0.91—0.96 для всех испытуемых. Следовательно, исследуемое изменение частоты основного тона характеризует интонационное выражение целой единицы, а не словесного ударения.

Итак, мы установили, что и в группе О—Д определение не может образовать самостоятельной интонационной единицы, и, следовательно, группа определение—дополнение безусловно является одной интонационной единицей и в предложениях типа П—С—О—Д образует конечную интонационную единицу.

Оказалось также, что интонационный контур этой конечной единицы может быть получен с помощью одного импульса увеличения частоты основного тона, начало которого совпадает с ударным слогом определения, а конец — с ударным слогом дополнения. Следовательно, элементом интонационной программы, соот-

ветствующим интонационной единице, является импульс увеличения частоты.

Далее мы установили, что интонационные контуры групп определение—дополнение в предложении типа П—С—О—Д и сказуемое—дополнение в предложениях типа П—С—Д совершенно одинаковы. Это позволяет нам утверждать, что группа сказуемое—дополнение также образует одну конечную интонационную единицу. Более того, эти данные подтверждают справедливость предположения, что в предложениях исследуемого типа конечная интонационная единица включает в себя два последних слова предложения. Этот вывод подтверждает и результаты предварительного анализа предложений типа *Мама — маляр*. Оказалось, что интонационные контуры таких предложений представляют собой одну интонационную единицу и притом — конечную.

Относительно неконечных единиц можно сказать, что в предложениях типа П—С—Д и О—П—С—Д их по одной и образованы они в первом случае подлежащим, и во втором — группой определение—подлежащее, а в предложении типа П—С—О—Д их две и образованы они подлежащим и сказуемым соответственно. Можно также утверждать, что в предложениях указанных типов неконечная интонационная единица выделяется импульсом увеличения частоты основного тона, начало которого совпадает с ударным слогом последнего слова в единице, а конец — с концом самой интонационной единицы.

В заключение следует отметить, что дальнейший анализ интонационных характеристик предложений других типов и выделение характерных особенностей получающихся интонационных контуров могут дать критерий для членения предложений на интонационные единицы (синтагмы) по виду интонационной кривой предложения. При этом необходимо, чтобы для членения использовался общий для всех типов предложений критерий выделения пограничных сигналов и получающиеся в результате такого членения синтагмы не противоречили лингвистическим критериям.

Л и т е р а т у р а

- Б р ы з г у н о в а Е. А. 1963. Практическая фонетика и интонации русского языка. М.
З и н д е р Л. Р. 1960. Общая фонетика. Л.
З л а т о у с т о в а Л. В. 1962. Фонетическая структура слова в потоке речи. Казань.
К р и в н о в а О. Ф. 1969. Фонетическая характеристика интонации одного типа простых предложений русского языка. Вестн. МГУ, 3, филол. : 63—77.
П а н о в М. В. 1968. Русская фонетика. М.
Ч и с т о в и ч Л. А., В. А. К о ж е в н и к о в, В. В. А л ь к р и н с к и й, Л. В. Б о н д а р к о, А. Г. Г о л у з и н а, Ю. А. К л а а с, Ю. И. К у з ь м и н, Д. М. Л и с е н к о, В. В. Л ю б л и н с к а я,

- Н. А. Федорова, В. С. Шупляков и Р. М. Шуплякова. 1965. Речь. Артикуляция и восприятие. М.—Л.
- Щерба Л. В. 1937. Фонетика французского языка. М.—Л.
- Bierwisch M. 1966. Regeln für die Intonation deutscher Sätze. *Studia grammatica*, VII : 99—201.
- Chomsky N., M. Halle, F. Lukoff. 1956. An Accent and juncture in English. In: For Roman Jakobson, the Hague : 65—80.
- Isačenko A. N., H. J. Schädlich. 1966. Untersuchungen über die deutsche Intonation. *Studia grammatica*, VII : 1—64.
- Karcevskij S. 1964. Sur la Phonologie de la Phrase. In: A Prague School Reader in Linguistics, Bloomington : 206—252.
- Lieberman Ph. 1967. Intonation, perception and language. Cambridge, Mass.
- Öhman E. S. 1967. Word and sentence intonation. A quantitative model. *STL—QPSR*, 2—3 : 20—57.
- Stockwell R. 1960. The place of intonation in a generative grammar of English. *Language*, 36, 3 : 360—371.
- Trager G. L., H. L. Smith. 1951. An outline of English structure. Ann. Arbor.
- Wells R. S. 1947. Immediate constituents. *Language*, 23, 2 : 83—117.

ИССЛЕДОВАНИЕ ВОСПРИЯТИЯ РУССКИХ СОГЛАСНЫХ МЕТОДОМ СЕРИЙНОГО ЗАПОМИНАНИЯ СЛОГОВ

В. И. Галунов, В. Ю. Шнейдер

Кафедра математической лингвистики ЛГУ

Процесс восприятия речи человеком можно представить в виде цепочки последовательных отображений входного стимула (Галунов, Слепокурова, 1971)

$$\omega_i \rightarrow x_i \rightarrow \xi_i \rightarrow R_i,$$

где ω_i — элементы множества входных сигналов, x_i — соответствующие множества первичных описаний (ощущений) входного сигнала в сенсорной системе, ξ_i — элементы множества описаний по субъективно полезным признакам, R_i — элементы множества выходных реакций. В рамках этой общей схемы попытаемся детализировать модель на последней ступени перехода $\xi_i \rightarrow R_i$. Представляется необходимым включить в это звено процедуру запоминания. При этом можно представить два варианта:

1) человек после выделения на некотором речевом сегменте субъективно полезного описания ξ_i запоминает это описание и далее при необходимости выдачи реакции R_i принимает решение о соответствующей R_i по запомненному описанию ξ_i и имеющейся дополнительной информации (например, о предшествующем или

последующем сегментах, некоторой априорной информации и т. п.);

2) человек на основании полученного субъективного полезного описания принимает решение о типе воспринятого речевого сегмента и далее запоминает это решение до момента необходимости выдачи решения R_i .

Оба варианта включения процедуры запоминания в модель восприятия речевых сигналов подразумевают использование так называемой кратковременной памяти (Broadbent, 1964).

Из сказанного предстает естественным при исследовании последней ступени восприятия речевых сигналов воспользоваться экспериментальными методами, принятыми при исследовании кратковременной памяти. В настоящей работе для исследования восприятия согласных был использован метод серийного запоминания слогов.

Общая схема эксперимента по методу серийного запоминания слогов выглядит следующим образом. Испытуемым предъявляются серии, состоящие из шести слогов. Длительность каждой серии примерно 4 сек. Интервал между сериями 8—10 сек. За это время испытуемый должен записать по памяти услышанную серию, причем запись он начинает только после окончания звучания данной серии.

Исследовалось запоминание следующих слогов: [pe], [fe], [be], [ve], [te], [se], [de], [ze]. Видно, что выбранные согласные (все они твердые) варьируются по трем признакам — озвонченности, способу и месту образования. Гласный [e] был выбран для того, чтобы избежать появления осмысленных звуко сочетаний во всей серии. Подбор слогов в серии осуществлялся по таблице случайных чисел с последующим отбрасыванием возникающих осмысленных сочетаний и кратных появлений одного или нескольких слогов.

В качестве испытуемых служили 20 человек в возрасте от 17 до 30 лет, с нормальным слухом. Прослушивание проводилось через телефоны по тракту связи, обеспечивающему стопроцентную разборчивость прослушиваемых таблиц. В каждом отдельном эксперименте испытуемым предъявлялось по 100 серий. Общее количество записанных испытуемыми по памяти серий 5200, т. е. 31 200 слогов. Такое общее количество слогов было подобрано экспериментально для обеспечения необходимой точности получаемых результатов. При меньшем количестве фиксируемых серий наблюдаются заметные флуктуации результатов.

При обработке результатов проводилось разграничение двух типов ошибок. Ошибками первого рода считались замены предъявленного согласного на другой; ошибками второго рода — перестановки двух или трех согласных в данной серии. Ввиду немногочисленности ошибок второго рода, они были исключены из дальнейшего рассмотрения. Ошибки первого рода далее приводятся в виде матриц ошибок.

Матрицы ошибок для исследуемых согласных даны в табл. 1 и 2. Числа в столбцах означают количество появлений вместо данного звука каждого из остальных. Количество правильных ответов получается вычитанием соответствующего числа в строке «Всего ошибок» из общего количества предъявлений данного звука. В табл. 1 результаты приведены в абсолютных величинах, а в табл. 2 результаты по каждому столбцу пересчитаны в про-

дентах. Средний процент ошибок 20.47%. Для различных испытуемых средний процент ошибок колебался от 5 до 60%.

При рассмотрении таблиц видно, что ряд по надежности запоминания согласных достаточно коррелирован (коэффициент ран-

Таблица 1

Полная матрица ошибок (в абс. вел.)

Воспроизведено	Передано								Всего
	<i>p</i>	<i>f</i>	<i>b</i>	<i>v</i>	<i>t</i>	<i>s</i>	<i>d</i>	<i>z</i>	
<i>p</i>		164	227	108	138	73	126	67	903
<i>f</i>	146		113	118	87	77	103	83	727
<i>b</i>	110	83		88	49	29	135	60	554
<i>v</i>	141	135	201		86	45	197	114	920
<i>t</i>	192	159	113	91		88	252	93	988
<i>s</i>	82	133	79	52	89		68	190	693
<i>d</i>	99	73	168	83	133	51		102	709
<i>z</i>	89	133	131	105	81	224	127		890
Всего ошибок . .	860	880	1032	645	663	587	1008	709	6384
Всего передано	3770	3744	3640	3900	3926	3770	4264	4186	31200

говой корреляции $R=0.80$) с рядом по частоте встречаемости этих согласных в речи. По данным Н. Д. Андреева (1967), вероятности встречаемости исследуемых согласных таковы: 0.019, 0.008, 0.11, 0.031, 0.039, 0.045, 0.016, 0.011.

В то же время ряд по вероятности перехода в соответствующий согласный других звуков практически не коррелирован с рядом по частоте встречаемости ($R=0.25$). Таким образом, полученные

Т а б л и ц а 2

Полная матрица ошибок (в %)

Воспроизведено	Передано								Всего
	<i>p</i>	<i>f</i>	<i>b</i>	<i>v</i>	<i>t</i>	<i>s</i>	<i>d</i>	<i>z</i>	
<i>p</i>	—	4.38	6.23	2.77	3.52	1.94	2.96	1.59	2.89
<i>f</i>	3.87	—	3.10	3.02	2.21	2.05	2.42	1.98	2.33
<i>b</i>	2.92	2.21	—	2.26	1.29	0.76	3.17	1.43	1.78
<i>v</i>	3.77	3.60	5.52	—	2.19	1.19	4.63	2.72	2.95
<i>t</i>	5.09	4.25	3.10	2.33	—	2.34	5.92	2.22	3.17
<i>s</i>	2.17	3.55	2.20	1.33	2.25	—	1.60	4.54	2.22
<i>d</i>	2.62	1.95	4.61	2.13	3.38	1.35	—	2.43	2.27
<i>z</i>	2.36	3.55	3.60	2.69	2.06	5.95	2.98	—	2.86
Всего ошибок . .	22.80	23.49	28.36	16.53	16.90	15.64	23.68	16.91	20.47
Всего передано	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

здесь результаты согласуются с результатами, полученными ранее для гласных (Галунов, Слепокурова, 1971).

В случае пересчета матрицы ошибок для групповых признаков звонкий—глухой, губной—переднеязычный и взрывной—щелевой (табл. 3—5) видно, что ранговые порядки вероятностей правильного узнавания признака и ранговые порядки перехода в данный признак совпадают с ранговыми порядками вероятностей встречаемости этих признаков в речи.

Таким образом, на основании изложенных в начале статьи соображений о возможном механизме участия кратковременной памяти в процессе восприятия речи можно предположить, что

Таблица 3

Матрица ошибок по признаку
звонкий—глухой
(в % от общего числа
предъявленных слогов)

Воспроизведе- дено	Передано		Всего
	звон- ные	глу- хие	
Звонкие . .	4.8	5.0	9.8
Глухие . .	6.1	4.6	10.7
Всего ошибок . .	10.9	9.6	20.5
Всего передано	48.8	51.2	100.0

Таблица 4

Матрица ошибок по признаку
губной—переднеязычный
(в % от общего числа слогов)

Воспроизведе- дено	Передано		Всего
	губ- ные	передне- язычные	
Губные . .	5.2	4.7	9.9
Передне- язычные	5.7	4.9	10.6
Всего ошибок . .	10.9	9.6	20.5
Всего передано	48.3	51.7	100.0

запоминаются не сами согласные фонемы, а их некоторые субъек-
тивные признаки, по всей вероятности весьма тесно связанные

Таблица 5

Матрица ошибок по признаку
взрывной—щелевой
(в % от общего числа слогов)

Воспроизведе- дено	Передано		Всего
	взрыв- ные	щеле- вые	
Взрывные	5.6	4.5	10.1
Щелевые	5.9	4.5	10.4
Всего ошибок . .	11.5	9.0	20.5
Всего передано	50.0	50.0	100.0

с артикуляторными признаками, рассматриваемыми в настоящей работе, т. е. следует предпочесть первый из предложенных выше вариантов восприятия.

Кроме того, при переходе от субъективно полезного описания воспринятых и запомненных сигналов к принятию решения, видимо, используется процедура типа Байесовой с учетом априорных вероятностей встречаемости отдельных признаков.

При принятии такой модели, конечно, могут быть объяснены только ранговые порядки надежности запоминания отдельных согласных и ранговые порядки их взаимных переходов. Отклонение же соответствующих абсолютных величин может являться след-

ствием действия целого ряда факторов: отклонения субъективных априорных вероятностей от истинных, неточного совпадения истинных субъективных признаков с принятыми нами артикуляторными и др.

Л и т е р а т у р а

- А н д р е е в Н. Д. 1967. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л.
- Г а л у н о в В. И., Н. А. С л е п о к у р о в а. 1971. Исследование восприятия русских гласных методом серийного запоминания слогов. *Вопр. языкознания* (в печати).
- B r o a d b e n t D. E. 1964. *Perception and communication*. N. Y.

ИССЛЕДОВАНИЕ СУБЪЕКТИВНОГО ПРЕДСТАВЛЕНИЯ ГРУППЫ РУССКИХ СОГЛАСНЫХ МЕТОДОМ СЕМАНТИЧЕСКИ ПРОТИВОПОЛОЖНЫХ ПАР

В. И. Галунов¹

Кафедра математической лингвистики ЛГУ

Одним из перспективных направлений в исследовании восприятия вообще и восприятия речи в частности является использование метода семантически противоположных пар (СПП), предложенного Осгудом (Osgood, 1952) для исследования восприятия смысловых отношений. При этом Осгуд предполагал обработку результатов типа фактор-анализа. Однако возможен и другой подход, опирающийся на теорию психологического шкалирования (Галунов, 1966).

Общая схема эксперимента по методу СПП выглядит следующим образом. Испытуемому задается некоторый конечный набор стимулов, восприятие которых предполагается исследовать, и набор признаков, задаваемых парами слов противоположного смысла, например: холодный—теплый, чужой—родной, шершавый—гладкий и т. п. От испытуемого требуется расположить исследуемые стимулы таким образом, чтобы самый «холодный» (для признака холодный—теплый) был слева, а самый «теплый» — справа, а остальные стимулы располагались в строчку между ними так, чтобы слева направо убывала «холодность» и возрастала «теплота». Такая процедура далее повторяется по всем заданным признакам. Возможны более простые и более сложные варианты этой экспериментальной методики, но в них во всех сохраняется принцип линейного описания исследуемых стимулов по признакам, заданным словами противоположного смысла. Здесь приведен вариант, который использовался нами в описанной далее работе.

При «шкальном» подходе данные, полученные по каждому из признаков, рассматриваются как одномерная шкала, сопоставленная исходному набору стимулов. При этом положение каждого

¹ В работе принимал участие студент Ю. Н. Храбрых.

из стимулов на этой шкале может быть получено как усредненный по отдельным опытам ранговый номер данного стимула. Набор линейных шкал в свою очередь может рассматриваться как система косоугольных координат в пространстве субъективно полезных признаков. Вычислив корреляции между отдельными признаками, мы можем определить число независимых субъективных признаков, а также значения координат каждого из стимулов в пространстве этих новых признаков и расстояния между стимулами в пространстве этих признаков.

Целью настоящей работы было выяснение возможности применения метода СПП для исследования восприятия согласных и получение содержательных результатов о структуре пространства субъективных признаков для исследуемой группы русских согласных.

В качестве признаков в настоящей работе были приняты следующие 9 пар слов:

- | | |
|------------------------|---------------------------|
| 1) гладкий — шершавый, | 6) чистый — нечистый, |
| 2) родной — чужой, | 7) приятный — неприятный, |
| 3) мягкий — жесткий, | 8) близкий — далекий, |
| 4) теплый — холодный, | 9) сладкий — горький. |
| 5) простой — сложный, | |

Эти пары были отобраны как дающие наиболее независимые результаты из принятой первоначально группы в 23 пары после предварительной проверки на небольшом числе испытуемых.

Исходный алфавит стимулов состоял из 8 открытых слогов, варьируемых только по своей согласной: *ка, та, па, ха, са, фа, ца, ша*. В качестве опорных были выбраны 6 первых согласных, варьируемых по способу (взрывной — щелевой) и месту (заднеязычный, переднеязычный, губной) образования. Слог *ца* был добавлен с целью рассмотрения положения аффрикат относительно групп взрывных и щелевых согласных. Слог *ша* интересен ввиду его особой артикуляции, представляемой различными фонетическими шкалами либо как двухфокусный, либо как переднеязычный небный и др.

Для признака мягкий — жесткий исследовалась еще группа слогов *ку, ну, пу, ху, су, фу, цу, шу* с целью выяснения возможного влияния контекстуального положения согласного на получаемые результаты.

Слоги предъявлялись зрительно. От испытуемых требовалось их внутреннее проговаривание в ходе эксперимента.

В качестве испытуемых использовались 12 студентов, не имеющих специальной фонетической подготовки.

Обработка результатов производилась следующим образом. Для каждого отдельного признака вычислялись коэффициенты ранговой корреляции по Спирмену (Ван-дер-Варден, 1960) между результатами всех испытуемых. Далее отбрасывались результаты тех испытуемых, которые оказались слабо коррелированными с остальными. Для оставшихся результатов по каждому из признаков строилась шкала, на которой стимулу i сопоставлялось число $\xi_i = \frac{\sum n_{ij}}{N}$,

где n_{ij} — порядковый номер i -го стимула j -го испытуемого; N — число испытуемых, по результатам которых проводилось усреднение. Далее вычислялась ранговая корреляция между признаками и расстояния между стимулами

$$d(i_1, i_2) = \sum_{k,l} (\xi_{i_1}^k - \xi_{i_2}^k) (\xi_{i_1}^l - \xi_{i_2}^l) \rho_{kl},$$

где k и l — индексы признаков, ρ_{kl} — корреляция между этими признаками.

Результаты шкалирования по первым 7 признакам приведены на рисунке. При этом по признаку мягкий—жесткий приведены усредненные результаты для слогов типа *Ca* и *Sy*. Такое усреднение было возможно, так как экспериментальные результаты по обоим типам слогов оказались одинаковыми, если рассматривать только порядок следования слогов. Отсев испытуемых, дающих ответы, не коррелированные с ответами основной группы испытуемых, для этих признаков составлял 1—2 человека. Дисперсия ответов в основной группе была меньше 1.6, а в среднем колебалась около 0.6. Для признаков близкий—далекий и сладкий—горький оказалось, что все испытуемые дают взаимно некоррелированные ответы. При дальнейшем рассмотрении эти признаки не учитывались. Причина отсутствия устойчивых результатов по этим признакам остается автору непонятной.

Для одного из испытуемых эксперимент был повторен трижды для признаков 1, 3, 6. Результаты, полученные в таких повторных экспериментах, оказались хорошо коррелированными ($R > 0.62$).

Рассчитанные на основании приведенных выше данных расстояния между согласными приведены в табл. 1. В табл. 2 приведены ранговые порядки близости между согласными. Верхняя строка в этой таблице указывает согласные, относительно которых отсчитываются расстояния, вторая строка — согласные, наиболее близкие к первым, третья — следующие за предыдущими, и т. д.

На основании всех приведенных результатов можно прежде всего сделать вывод о применимости метода СПП для исследования субъективных представлений согласных. Полученные данные носят явно систематический характер для большинства признаков. Отсутствие устойчивых ответов в отдельных случаях требует дополнительного исследования.

Далее, как показывает рассмотрение рисунка, конструируемые на основании пар слов семантически противоположного значения признаки в пространстве субъективных описаний не могут быть прямым образом сопоставлены с известными системами классификации согласных. В то же время при рассмотрении табл. 1, 2 можно заметить отчетливую закономерность. Наблюдается явная группировка взрывных согласных *п, т, к* (среднее расстояние между ними оказалось равным 2.7) и щелевых однофокусных согласных *ф, с, х* (расстояние между которыми было равно 7.0). При этом *с* занимает положение как бы промежуточное между взрывными и щелевыми. К группе *ф, с, х* явным образом примыкает *ш*. Однако расстояние между взрывными и *ш* явно больше, чем между взрывными и щелевыми. В первом приближении взаимное расположение групп взрывных, щелевых однофокусных и *ш* может быть представлено в виде линейной цепочки: *п, т, к—с—ф, х—ш*.

Следует иметь в виду, что при построении геометрической конфигурации на основании данных табл. 1 удастся построить

ξ_{i_1}	0	1	2	3	4	5	6	7	8
			ПТФ	К		С	ХШЦ		
			2.3	2.4	2.6		6.5	6.7	
ξ_{i_2}	0	1	2	3	4	5	6	7	8
			П	КТ		С	Ш	ФЦХ	
			1.6	2.5	2.7	4.8	5.3	6.5	6.6
ξ_{i_3}	0	1	2	3	4	5	6	7	8
			Ш	Ф	Х	С	П	Т	КЦ
			2.3				5.9	6.6	6.8
ξ_{i_4}	0	1	2	3	4	5	6	7	8
			Ф	Ш	ПХТ	К	Ц	С	
			2.3	3.3	3.6	5.5	6.4	6.6	
ξ_{i_5}	0	1	2	3	4	5	6	7	8
			С	ПТК		ФХ	Ц	Ш	
			2.4	2.7	2.8	5.5	5.6	6.7	7.4
ξ_{i_6}	0	1	2	3	4	5	6	7	8
			Т	П	К	С	Ц	Ш	Х
			2.6	2.9	3.5	5.5	6.4		
ξ_{i_7}	0	1	2	3	4	5	6	7	8
			ФК	Т		СХ	Ц	Ш	
			2.4	2.5	3.2	5.7	6.3	6.8	
						5.8			

Результаты шкалирования согласных по отдельным признакам.

Таблица 1

Расстояния между отдельными
согласными

	п	т	к	ф	с	х	ш
т	2.1						
к	3.5	2.4					
ф	8.9	9.6	10.3				
с	6.8	6.4	6.4	8.5			
х	10.5	10.4	10.8	5.9	6.6		
ш	11.5	11.6	12.1	6.8	8.0	7.1	
ц	10.9	9.8	9.0	10.7	7.2	14.0	8.4

Таблица 2

Ранговые порядки близости
между согласными

	п	т	к	ф	с	х	ш	ц
1	т	п	т	х	т	ф	ф	с
2	к	к	п	ш	к	с	х	ш
3	с	с	с	с	х	ш	с	к
4	ф	ф	ц	п	п	т	п	т
5	х	ц	ф	т	ц	п	т	ф
6	ц	х	х	к	ш	к	к	п
7	ш	ш	ш	ц	ф	ц	ц	х

только не менее чем трехмерную модель для группы *п, т, к, ф, с, х, ш*. При этом не удается найти отчетливых объективных коррелят для субъективных осей, кроме оси взрывные—целевые. Априори можно было предполагать, что субъективно аффрикаты являются промежуточным типом между взрывными и целевыми. Рассмотрение приведенных выше результатов показывает, что это не так. Согласный *ц* оказывается равно далеким и от взрывных, и от целевых, притом расстояние до этих групп существенным образом превышает расстояние, допустимое для занятия места целевым и взрывным. Таким образом, следует вывод, что аффрикативный характер согласного соответствует некоторому особому признаку.

Литература

- Г а л у н о в В. И. 1966. Психофизические шкалы. В сб.: Автоматическое распознавание слуховых образов, Новосибирск.
В а н - д е р - В а р д е н Б. Л. 1960. Математическая статистика. М.
O s g o o d С. E. 1952. The nature and measurement of meaning. Psychol. Bull., 49 : 197.

ВОСПРИЯТИЕ ФОНЕМ В СЛОГАХ РАЗЛИЧНЫХ ТИПОВ

*Л. В. Бондарко, Л. Е. Буколыщикова, Л. П. Павлова,
Н. Д. Светозарова, А. С. Штерн*

Лаборатория экспериментальной фонетики им. Л. В. Щербы, ЛГУ

При описании инвариантных признаков фонем пытаются обычно найти те акустические или артикуляционные характеристики, которые связаны с отдельными звуками речевого потока (Якобсон, Халле, 1962). Однако это едва ли возможно, так как ни с точки зрения механизмов образования речевых последовательностей, ни с точки зрения известных сейчас процессов восприятия минимальный сегмент не соответствует фонеме, а совпадает с сочетанием фонем типа CV (Чистович и др., 1965; Бондарко, 1969). В связи с этим изучение минимальных единиц типа открытого слога приобретает большое значение и при описании акустико-артикуляторных изменений, происходящих в речевом потоке, и при выявлении специфических закономерностей восприятия звуков речи (Бондарко и др., 1966б).

Настоящая работа посвящена исследованию зависимости восприятия слогов типа CV, находящихся в разных положениях по отношению к ударению и представляющих сочетания различных фонем. Особенный интерес в этом отношении представляет изучение восприятия безударных слогов, которые, как известно, составляют в речевом потоке абсолютное большинство. Относительно русского языка систематических данных подобного рода нет. Исследования, посвященные восприятию сегментов, соответствующих безударным гласным и согласным (Вербицкая, 1965; Бондарко и др., 1966б), показывают, что в большинстве случаев ответы аудиторов являются случайными, что связано с малой длительностью безударных звуков и их значительными качественными изменениями по сравнению с ударными. В этом отношении слоги CV должны представлять более удобный материал, так как и временные характеристики слога больше соответствуют постоянной времени человеческого уха, да и качественные изменения звуков в пределах слога являются более предсказуемыми с точки зрения их возможных взаимовлияний в речевом потоке.

Задачей данной работы является исследование того, насколько существенно влияет на восприятие слога его место по отношению к ударению и его фонетическая структура, т. е. качество согласного и гласного, образующих слог.

Относительно влияния места ударения на степень изменения безударного слога в фонетической литературе существуют два различных представления (речь идет о русском языке). Наиболее распространенным и традиционным является мнение, согласно которому характер изменения безударного слога не зависит от

качества компонентов этого слога, т. е. от качества согласного и гласного. С этой точки зрения меньше всего изменяется слог, непосредственно предшествующий ударному (первый предударный); второй предударный слог и заударный слог изменяются в значительно большей степени и в этом отношении противопоставлены как ударному, так и первому предударному слогу (Потебня, 1865; Аванесов, 1956). Другое представление, основанное в отличие от первого на анализе экспериментальных данных, заключается в том, что на характер редукции (изменения безударного слога по сравнению с ударным) влияет не только место ударения, но и качество гласного: слоги с гласным [a] действительно можно сгруппировать так, как предлагал еще А. А. Потебня, а слоги с закрытыми гласными [u], [i], [ɨ] характеризуются специфическими отношениями: во всяком случае, если рассматривать изменение этих гласных в безударных слогах, то ни по временным, ни по спектральным характеристикам они не разбиваются на две группы, как это происходит в слогах с гласным [a] (Бондарко и др., 1966а).

Нужно, однако, заметить, что систематического исследования восприятия безударных и ударных слогов на материале русского языка до сих пор не было произведено, и мы практически не знаем, каким образом оценивает ритмическую структуру слова человек, анализирующий речевой поток.

МАТЕРИАЛ И МЕТОДИКА

В качестве экспериментального материала были выбраны слоги с переднеязычными согласными и гласными [a], [u], [i], [ɨ]. Выбор переднеязычных согласных обусловлен тем, что именно в этой группе представлены самые разнообразнейшие типы согласных — шумные и сонанты, чистые и носовые, плавные и дрожащие, взрывные и аффрикаты, однофокусные и двухфокусные и т. д. Выбор гласных [a], [u], [i], [ɨ] обусловлен тем, что именно эти гласные встречаются как в ударных, так и в безударных слогах, что очень существенно с точки зрения тех задач, которые поставлены в данной работе.

Каждый слог находился в следующих фонетических позициях.

1. Являлся ударным слогом слова (например, *сухо*).
2. Являлся первым предударным слогом (*сух*а).
3. Являлся вторым предударным слогом (*сух*оват).
4. Являлся заударным слогом (*насух*о).

Ниже слоги обозначаются следующим образом: указывается начальный согласный (или символ С, обозначающий любой согласный), гласный и индекс, соответствующий номеру одной из перечисленных позиций слога по отношению к ударению ([Ca]₁ — ударный слог с любым переднеязычным твердым согласным и гласным [a]; [sa]₁ — ударный слог с начальным глухим целевым [s] и гласным [a]).

Дикторами были мужчина 30 лет и женщина 35 лет, оба преподаватели русской фонетики. Мужчина — представитель московской произносительной нормы, женщина — ленинградской.

Слова читались по списку, с небольшими паузами между ними, в нормальном темпе. Запись производилась на магнитофоне МЭЗ 28-А со скоростью движения ленты 381 мм/сек. Затем из записанных слов были выделены

при помощи сепаратора интересующие нас слоги (по двести слогов в произнесении каждого диктора).¹

Выделенные слоги предъявлялись аудиторам в двух различных опытах. В первом опыте перед аудиторами ставилась задача определить согласный и гласный в слоге. Для этого опыта все выделенные слоги были записаны на магнитную ленту в случайном порядке и разделены на 4 серии по 50 слогов в каждой. Каждый слог предъявлялся каждому аудитору один раз. Группа аудиторов состояла из 25 человек (студенты I—III курсов филологического факультета). Испытуемые использовали для записи буквы русского алфавита и иногда для обозначения гласного употребляли *ъ*, т. е. знак, употребляемый обычно для обозначения гласного заднего ряда, сильно редуцированного. Полученные данные обрабатывались следующим образом: определялось количество правильно опознанных слогов (в % к общему числу предъявлявшихся в опытах слогов). Все слоги были при этом разбиты на 5 групп с учетом твердости или мягкости согласного: С [а], С [u], С' [u], С [i], С' [i].

Во втором опыте перед аудиторами ставилась другая задача: они должны были отмечать в анкетах, какие они слышат слоги — одинаковые или разные. При этом серии состояли из слогов, записанных на магнитную ленту попарно: CV₁—CV₂; CV₃—CV₄; CV₂—CV₃. Аудиторы знаком «плюс» отмечали, что два слога звучат одинаково, знаком «минус» обозначали разные, с их точки зрения, слоги.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Анализ ответов аудиторов, полученных в первом опыте, показал, что опознавание слога существенно меняется в зависимости

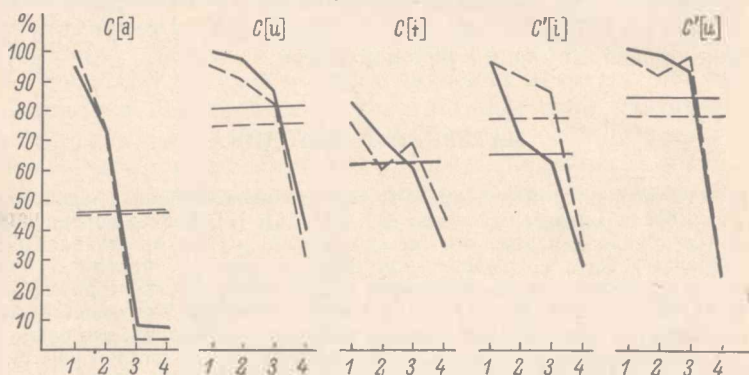


Рис. 1. Опознаваемость слогов в зависимости от качества гласного и места слога по отношению к ударению.

По оси абсцисс — типы слогов: 1 — ударный, 2 — первый предударный, 3 — второй предударный, 4 — заударный; по оси ординат — процент правильных ответов. Сплошная линия — диктор С., прерывистая — диктор Б. Линиями, параллельными оси абсцисс, обозначена средняя опознаваемость слога (независимо от места ударения).

от позиции: лучше всего опознаются ударные и первые предударные слоги, хуже всего — заударные. На рис. 1 приведены данные об опознаваемости слогов, сгруппированных по позициям по

¹ Число возможных слогов, начинающихся с переднеязычных, значительно больше, но в описываемом материале не представлены слоги с начальными звонкими и аффрикатами, так как по характеру контрастирования они примыкают к другим группам переднеязычных.

отношению к ударению. На рисунке достаточно хорошо видно, что и качество гласного влияет на характер опознавания: если ударные слоги с гласными [a], [u], [i] опознаются в 90—100% случаев, то слоги с гласными [ɪ] даже под ударением опознаются лишь в 75—80% случаев. Анализ анкет аудиторов показывает, что такое ухудшение распознавания [ɪ]-слогов происходит, как правило, за счет гласного: 10—15% ударных слогов с гласным [ɪ] опознаются как слоги с любым другим гласным. Согласный при этом опознается правильно.

Влияние места ударения на опознаваемость безударного слога также различно в зависимости от качества гласного. Слоги с гласным [a] составляют отдельную группу, характеризующуюся тем, что максимально низкий процент опознаваемости наблюдается для второго предударного и заударного слогов. В остальных случаях второй предударный опознается значительно лучше, чем заударный, и в этом отношении он гораздо ближе к первому предударному.

Рассмотрим более детально, каковы ошибки аудиторов при восприятии слогов. Испытуемые в своих анкетах обозначали предъявлявшиеся слоги следующим образом:

а) правильно (т. е. с правильными обозначениями и согласного и гласного);

б) как один согласный (тот, который действительно имелся в предъявляемом слоге);

в) как один гласный (также тот, который содержится в слоге);

г) как согласный и какой-либо другой гласный (т. е. не тот, который реально содержался в слоге).

На рис. 2 эти данные приведены для [a]-слогов, [u]-слогов и [i]/[ɪ]-слогов (независимо от твердости—мягкости начального

Таблица 1

Опознаваемость слога как одного гласного
(в % к числу ответов)

Начальный согласный	Воспринят как один гласный	
	во втором предударном слог	в заударном слог
Диктор С.		
Шумный	0	0
Сонант	5	26
Диктор Б.		
Шумный	0	0
Сонант	6	29

Опознаваемость слогов с переднеязычными согласными и гласными

Слог	Позиция слога	Процент случаев правильного опознавания		Слог	Позиция слога	Процент случаев правильного опознавания	
		Диктор Б.	Диктор С.			Диктор Б.	Диктор С.
[ta]	1	100	100	[nu]	1	91	100
	2	100	100		2	96	88
	3	8	0		3	40	100
	4	0	0		4	13.5	33
[sa]	1	100	86.5	[lu]	1	100	100
	2	64	70		2	100	92
	3	4	4		3	88	92
	4	4	0		4	19	72
[sa]	1	100	100	[l'u]	1	96	100
	2	83.5	52		2	100	92
	3	4	4		3	100	87.5
	4	4.5	0		4	0	9
[na]	1	100	96	[ru]	1	96	100
	2	0	60		2	100	100
	3	0	29		3	79	56.5
	4	5	33		4	0	4
[la]	1	100	100	[r'u]	1	92	100
	2	88	75		2	88	96
	3	0	0		3	88	87
	4	4	0		4	0	33
[ra]	1	100	100	[ti]	1	82	76
	2	100	96		2	35	76
	3	0	12.5		3	65	30
	4	5	0		4	60	24
[tu]	1	100	100	[si]	1	100	88
	2	100	100		2	79	92
	3	100	100		3	88	56.5
	4	52	72		4	37.5	16
[t'u]	1	100	100	[ni]	1	67	64
	2	72	96		2	48	54
	3	100	100		3	—	—
	4	64	76		4	17	36
[su]	1	100	100	[li]	1	33	100
	2	100	100		2	56	88
	3	96	100		3	—	88
	4	87.5	52		4	—	76
[s'u]	2	—	100	[ri]	1	96	72
	3	—	100		2	76	32
[šu]	1	96	100		3	65	76
	2	46	100		4	32	5
	3	80	80				
	4	16.5	9				

Таблица 2 (продолжение)

Слог	Позиция слога	Процент случаев правильного опознавания		Слог	Позиция слога	Процент случаев правильного опознавания	
		Диктор Б.	Диктор С.			Диктор Б.	Диктор С.
[t'i]	1	96	100	[l'i]	3	75	44
	2	40	54		4	63	32
	3	44	67		1	92	96
	4	20	52		2	73	96
[s'i]	1	100	96	[r'i]	3	18	100
	2	48	80		4	18	35
	3	52	72		1	96	96
	4	25	18		2	27	48
[n'i]	1	92	100		3	0	40
	2	83	96		4	27	7

согласного). Как можно увидеть на этом рисунке, во всех случаях ухудшение распознавания происходит за счет гласного, а не за счет согласного в слоге: так, второй предударный и заударный слоги с гласным [a] опознаются правильно лишь в 10% случаев, но согласный опознается в 90% случаев; для слогов с гласными [u], [i], [ɨ], при низкой опознаваемости в заударной позиции, также характерно, что согласный опознается правильно достаточно часто.

Случаи, когда слог опознавался как один гласный, относительно редки, однако важно отметить, что большая их доля падает, во-первых, на слоги с начальными сонантами, а во-вторых, на заударные позиции (табл. 1).

Нужно еще заметить, что случаи, когда аудиторы отказывались вообще опознать слог, сравнительно редки при предъявлении ударных и предударных слогов и более часты при предъявлении заударных.

Обобщая данные, приведенные на рис. 2, можно сказать, что и с точки зрения распределения правильных и ошибочных ответов слоги с гласным [a] противопоставлены слогам с гласными [u], [i], [ɨ], так как [a]-слоги разделяются на ударные и первые предударные, с одной стороны, и вторые предударные и заударные — с другой; для остальных слогов распределение ответов при предъявлении ударных и предударных слогов в принципе одно и то же, тогда как заударные в этом отношении образуют отдельную группу.

Для более полного представления о характере опознавания слогов приводится табл. 2, в которой даются сведения о правильном опознавании каждого из предъявленных слогов.

Рассмотрим результаты опытов, в которых проводилась парная идентификация слогов. На рис. 3 приведены данные о числе случаев идентификации пар слогов CV_1-CV_2 и CV_3-CV_4 .

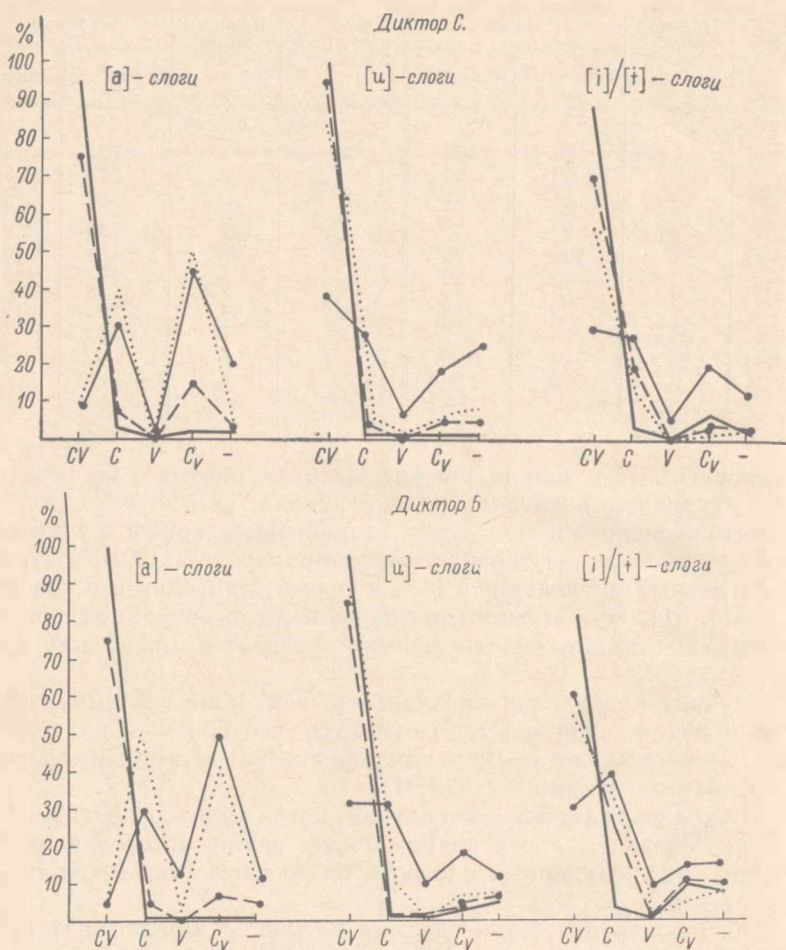


Рис. 2. Характер ошибок при восприятии слогов (отдельно для каждого диктора).

По оси абсцисс — ответы аудиторов: CV — правильное опознание слога, C — опознание слога как одного согласного, V — опознание слога как одного гласного, CV — опознание слога как согласного с другим гласным (не содержащимся в слоге), черточка — отказ; по оси ординат — количество ответов в %. Сплошная жирная линия — ударные слоги, прерывистая линия — первые предударные слоги, точечная линия — вторые предударные слоги, сплошная тонкая линия — заударные слоги.

Прежде всего обращает на себя внимание тот факт, что первый предударный и ударный слоги очень часто опознаются аудиторами как одинаковые (в среднем в 80—95% случаев, за исключением

[а]-слогов, о которых речь будет идти ниже). Известно, что наиболее общим и регулярным различием между ударными и безударными является различие по длительности: первый предупредный слог в среднем в полтора раза короче ударного слога (Бондарко и др., 1966а). Однако аудиторы обозначают слоги CV_1 и CV_2 как

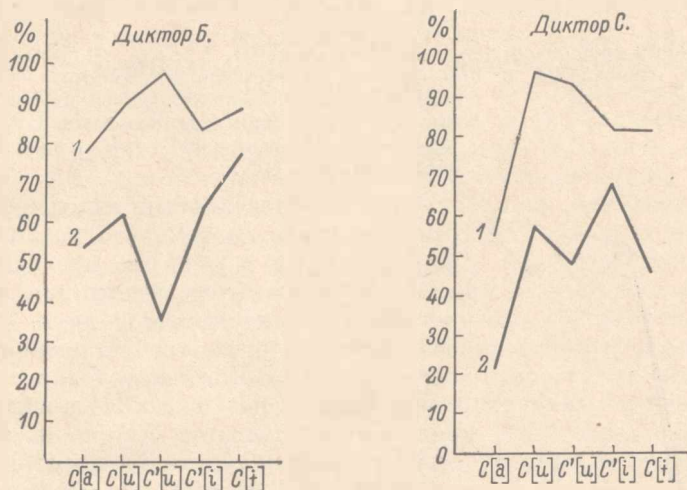


Рис. 3. Идентификация слогов CV_1 — CV_2 и CV_3 — CV_4 .

По оси абсцисс — типы слогов; по оси ординат — процент случаев идентификации. 1 — идентификация слогов CV_1 — CV_2 ; 2 — идентификация слогов CV_3 — CV_4 .

одинаковые, отвлекаясь от этого различия, хотя слоги следуют непосредственно друг за другом.

При идентификации слогов с CV_3 — CV_4 наблюдается принципиально иная картина, так как эти слоги воспринимаются как одинаковые гораздо реже, чем слоги CV_1 — CV_2 . Это различие является существенным и подтверждается статистически.

Приведем оценку расхождения в двух сериях для одного диктора:

I серия	II серия
$\bar{x}_1 = 22.2$	$\bar{x}_2 = 11.8$
$s_1 = 2.05$	$s_2 = 3.66$
$\sigma_1 = 2.15$	$\sigma_2 = 3.81$
$n_1 = 12$	$n_2 = 13$

Нулевая гипотеза о равенстве средних проверяется по t -критерию. Применение этого критерия предполагает равенство дисперсий. Проверим вначале это предположение по критерию F Фишера.

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{3.81^2}{2.15^2} = 3.14.$$

При 12 и 11 степенях свободы при уровне значимости 0.05 $F_{\text{табл}} = 8.88$. Так как $F < F_{\text{табл}}$, дисперсии равны. Теперь проверим гипотезу о равенстве средних.

$$t = \frac{22.2 - 11.8}{\sqrt{\frac{12 \cdot 2.05^2 + 13 \cdot 3.66^2}{12 + 13 - 2} \cdot \left(\frac{1}{12} + \frac{1}{13}\right)}} = 8.32.$$

Для степени свободы $n_1 + n_2 - 2 = 23$ $t_{\text{табл}} = 2.07$, $t > t_{\text{табл}}$. Поэтому нулевая гипотеза отвергается и принимается альтернативная гипотеза — среднее первой выборки больше среднего второй выборки, т. е. идентифицирование слогов CV_1 и CV_2 происходит гораздо чаще, чем идентификация слогов CV_3 и CV_4 .

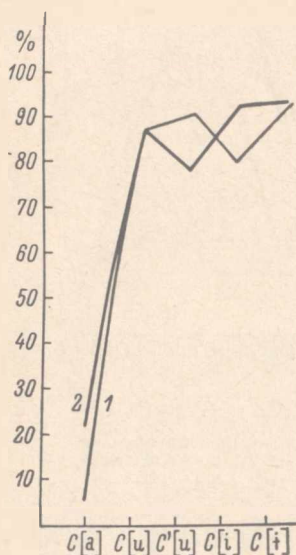


Рис. 4. Идентификация слогов $CV_2 - CV_3$.

По оси абсцисс — типы слогов; по оси ординат — процент случаев идентификации. 1 — диктор Б.; 2 — диктор С.

Этот парадоксальный вывод, который свидетельствует о том, что различия между ударными и первыми предударными слогами менее существенны для аудиторов, чем различия между двумя видами безударных слогов, требует специального рассмотрения. Для выявления причин столь резкого расхождения данных первого и второго опытов был поставлен третий, в котором аудиторам предъявлялись пары слогов $CV_2 - CV_3$, т. е. первый и второй предударные. На рис. 4 представлены эти данные. Анализ результатов этого опыта показывает, что [a]-слоги занимают совершенно исключительное место: они опознаются как одинаковые менее чем в 10% случаев, т. е. аудиторы совершенно четко различают первый и второй предударные [a]-слоги по их собственным характеристикам. Все остальные слоги (т. е. слоги с гласными [u], [i], [i] не различаются аудитором: первый и второй предударные слоги в 80% и выше опознаются как одинаковые.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные данные позволяют сделать вывод о том, что с точки зрения восприятия существенным является не только место слога по отношению к ударению, но и фонетический состав слога (и в первую очередь — качество гласного). Оказывается, что только характеристики слогов с гласным [a] подтверждают традиционную гипотезу о существовании двух степеней редукции в безударном положении: действительно, [a]-слоги по некоторым своим признакам разбиваются на 3 группы — ударные, первые пред-

ударные (1-я степень редукции) и вторые предударные и заударные (2-я степень редукции). Данные восприятия подтверждают это следующим образом: при стопроцентном распознавании ударных [а]-слов первые предударные распознаются в 70% случаев, а вторые предударные и заударные — менее чем в 10% случаев. Данные по идентификации слогов также подтверждают этот факт: ударные и первые предударные [а]-слоги опознаются как одинаковые гораздо реже, чем все остальные слоги; при идентификации безударных слогов [а]-слоги также различаются больше, чем все другие слоги. Существенно, однако, заметить, что с гипотезой о двух степенях редукции безударных плохо согласуется тот факт, что вторые предударные и заударные [а]-слоги, которые с точки зрения этой гипотезы должны образовывать одну группу, в описанных выше опытах идентифицировались достаточно редко (в 52% случаев у одного диктора и в 22% случаев у другого).

Относительно слогов с гласными [u], [i], [i] можно утверждать, что предположение о двух степенях редукции безударных в этих случаях не подтверждается данными восприятия. Здесь можно говорить лишь о противопоставлении ударных и предударных слогов заударным. Этот факт иллюстрирует как данные об опознаваемости слогов и распределении правильных и неправильных ответов при их опознавании, так и результаты опытов по идентификации слогов: оказывается, что ударные и предударные слоги отличаются друг от друга в меньшей степени, чем предударные от заударных. Однако нужно заметить, что последнее утверждение справедливо, видимо, лишь для заударных слогов, выделенных из изолированно произнесенных слов; в более длинных речевых последовательностях (словосочетаниях и предложениях) такого резкого противопоставления предударных заударным не наблюдается ни в отношении их акустических характеристик, ни в отношении различий при их восприятии (Бондарко и др., 1968).

Безударные слоги опознаются хуже, чем ударные независимо от качества гласного в слогe. Ухудшение же распознавания происходит, как правило, не за счет согласного, а за счет гласного, который может или вообще не опознаваться аудиторами, или восприниматься как другой гласный. Хуже всего всегда опознаются заударные слоги, которые и в этом отношении противопоставлены всем остальным слогам.

ВЫВОДЫ

1. Распознавание фонем в пределах открытого слога зависит от двух факторов: во-первых, от положения данного слога по отношению к ударению; во-вторых, от качества гласного в слогe.

2. По сравнению с распознаванием выделенных из речевой цепи звуковых сегментов, соответствующих фонемам, распознава-

ние в пределах слога дает лучшие результаты, что особенно заметно для согласных. Гласные в ряде случаев (в безударных слогах) или вообще не воспринимаются, или опознаются неверно.

3. Наиболее общим для всех слогов с точки зрения восприятия является противопоставление ударных и предударных слогов заударным.

4. Характеристики предударных слогов с гласными [u], [i], [i] не являются достаточными для того, чтобы носители языка могли с уверенностью определять их место в слове по отношению к ударению. Данные опытов по восприятию свидетельствуют о том, что при определении ритмической структуры слова недостаточно сведений об отдельных слогах, входящих в анализируемое слово, и что для определения места ударного слога нужны сравнительные характеристики всей последовательности слогов.

Л и т е р а т у р а

- А в а н е с о в Р. И. 1956. Фонетика современного русского литературного языка. М.
- Б о н д а р к о Л. В. 1969. Слоговая структура речи и дифференциальные признаки фонем. Автореф. докт. дисс. Л.
- Б о н д а р к о Л. В., Л. А. В е р б и ц к а я, Л. Р. З и н д е р. 1966а. Акустические характеристики безударности. В сб.: Структурная типология языков, М.: 56—64.
- Б о н д а р к о Л. В., Л. А. В е р б и ц к а я, Л. Р. З и н д е р, Л. П. П а в л о в а. 1966б. Различаемые звуковые единицы русской речи. В сб.: Механизмы речеобразования и восприятия сложных звуков, М.—Л.
- Б о н д а р к о Л. В., Л. Р. З и н д е р, Н. Д. С в е т о з а р о в а. 1968. Разграничение слов в потоке речи. *Вопр. языкознания*, 2: 68—81.
- В е р б и ц к а я Л. А. 1965. Звуковые единицы русской речи и их соотношение с оттенками и фонемами. Автореф. канд. дисс. Л.
- П о т е б н я А. А. 1865. О звуковых особенностях русских наречий. *Филол. записки*, вып. 1.
- Ч и с т о в и ч Л. А., В. А. К о ж е в н и к о в, В. В. А л я к р и н с к и й, Л. В. Б о н д а р к о, А. Г. Г о л у з и н а, Ю. А. К л а а с, Ю. И. К у з ь м и н, Д. М. Л и с е н к о, В. В. Л ю б л и н с к а я, Н. А. Ф е д о р о в а, В. С. Ш у п л я к о в, Р. М. Ш у п л я к о в а. 1965. Речь. Артикуляция и восприятие. Л.—М.
- Я к о б с о н Р., М. Х а л л е. 1962. Фонология и ее отношение к фонетике. В сб.: Новое в лингвистике, 2, М.

ИССЛЕДОВАНИЕ МЕХАНИЗМА РЕГУЛИРОВАНИЯ ИНТЕНСИВНОСТИ ШУМНЫХ СОГЛАСНЫХ

Т. Г. Малиникова, А. В. Венцов

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

При исследовании восприятия громкости шумных согласных методом имитации (Малиникова, 1970) было установлено, что уровень интенсивности ответа испытуемого связан с уровнем интенсивности стимула линейной зависимостью, причем значения уровней интенсивности ответа и стимула не совпадают по величине. Если в свете этих данных исследовать механизм регулирования интенсивности шумных согласных, задавшись целью определить управляемые элементы многоступенчатой системы создания таких звуков, то естественно предположить следующее. При восприятии и воспроизведении интенсивности не используется акустическая обратная связь и не регулируется непосредственно интенсивность выходного сигнала, т. е. управление не производится таким способом, как это представлено на рис. 1, *I*, какого можно было бы ожидать на основании гипотезы Бернштейна об организации движений (Бернштейн, 1947). Регулируемым же элементом является комплекс физиологических параметров речеобразующего тракта. Схема управления такова: в ответ на интенсивность стимула в нервной системе испытуемого создается набор команд к тем органам, которые определяют необходимый комплекс физиологических параметров, при этом решается задача — обеспечить требуемую величину параметров. Интенсивность акустического выходного сигнала, которую создает испытуемый, определяется только величиной этих управляемых параметров и влиянием остальных органов речеобразующего тракта (рис. 1, *II*).

Эксперименты, данные которых представлены в упомянутой выше работе по имитации, были достаточно сложными, и они не позволяют непосредственно считать наше предположение утверждением, для этого потребовалась дополнительная проверка, которой и посвящена данная статья.

Был использован метод искусственного изменения одного из параметров системы речеобразования, влияющего на интенсивность акустического сигнала, создаваемого испытуемым. Если наше предположение верно, то изменения условий образования

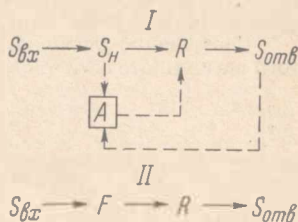


Рис. 1. Схематическое изображение механизма управления интенсивностью.

I — первый способ ($S_{вх}$ — интенсивность входного сигнала, S_n — интенсивность намерения, $S_{отв}$ — интенсивность выходного сигнала, A — блок сравнения, R — реализация команд); *II* — второй способ (F — набор команд; остальные обозначения те же, что в *I*).

ответного акустического сигнала не должны влиять на величины регулируемых физиологических параметров, но должны значительно повлиять на значения интенсивностей ответов. А если все-таки испытуемый регулирует непосредственно интенсивность, то такое искусственное вмешательство не должно влиять на результаты, и ответы должны быть одинаковы в нормальных и измененных условиях.

МЕТОДИКА

Как известно, акустический шум, являющийся одним из основных признаков шумных щелевых согласных, образуется при прохождении воздушной струи через сужение в определенном месте речевого тракта. При этом интенсивность шума зависит от размеров и формы образующейся в речевом тракте щели, а также от величины давления воздуха за местом образования щели (Meyer-Eppler, 1953; Fant, 1960).

Совершенно очевидно, что при многократном произнесении одного и того же шумного согласного размеры и форма образующейся щели изменяются

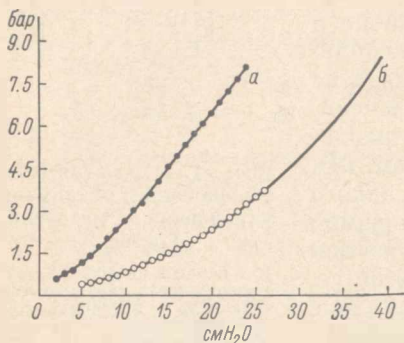


Рис. 2. Зависимость звукового давления на выходе речевого тракта (а) и дюзы (б) (по оси ординат, бар) от давления воздуха за щелью в тракте или за дюзой (по оси абсцисс, см вод. ст.).

сечения шумного согласного через систему с обратной связью регулируется непосредственно интенсивность акустического сигнала.

Для проведения исследования была изготовлена дюза в виде пластмассовой пластинки с овальным отверстием размерами 3×10 мм. Для увеличения интенсивности генерируемого дюзой шума у выходного ее отверстия был укреплен щиток, о который разбивалась струя выходящего воздуха. Для сравнения характеристик дюзы и естественной щели в речевом тракте были исследованы зависимости звукового давления на выходе щели и дюзы от давления воздуха за щелью и дюзой (рис. 2). Оказалось, что изготовленная нами дюза по своим характеристикам значительно уступает естественной щели в речевом тракте.

Исследование проводилось методом имитации, так как необходимость имитировать стимулы разной интенсивности заставляет испытуемого создавать сигналы, изменяющиеся с заданным шагом в определенном диапазоне интен-

от произнесения к произнесению, это при прочих равных условиях должно приводить к значительному разбросу величин интенсивности произносимого согласного, если эта интенсивность непосредственно не регулируется. Если теперь естественную щель в речевом тракте заменить искусственной и заставить испытуемого создавать подобный заданному шумному согласному звук, продувая воздух через дюзу постоянных размеров, разброс интенсивностей многократно произнесенного в новых условиях «согласного», при прочих равных условиях, должен уменьшиться. Возможно, изменится и абсолютная величина интенсивности создаваемых таким способом звуков, так как к. п. д. дюзы, как генератора шума, может отличаться от к. п. д. естественной щели в речевом тракте.

Описанные выше изменения интенсивности должны отсутствовать при условии, что во время произне-

сивностей. Во всех опытах имитировалась интенсивность синтетического шумного согласного [ʃ].

Для этого набор из десяти стимулов разной интенсивности был синтезирован и записан на магнитную пленку в таком порядке, что каждый стимул предъявлялся по одному разу после всех остальных и один раз после самого себя. Всего было записано 102 стимула. Блок-схема установки для синтеза и записи стимулов на магнитную пленку приведена на рис. 3.

На пленке каждому стимулу предшествует маркирующий сигнал (3в, 18 кгц), используемый для запуска автоматической системы измерения и регистрации интенсивностей стимула и ответа испытуемого во время опыта. Схема временной организации маркирующего сигнала и стимула показана на рис. 4.

Интенсивность стимулов изменялась ступенями по 3.47 дБ от уровня самого громкого стимула, эффективное напряжение которого при записи на магнитную пленку было равно 3 в. При десяти стимулах полный диапазон изменения их интенсивности составляет 31.25 дБ.

Опыты проводились в звукозаглушенной камере. Стимулы подавались на головные телефоны, при этом было предусмотрено, чтобы стимулы с максимальной интенсивностью не вызывали болевых ощущений у испытуемого и могли быть симитированы без напряжения. Блок-схема установки для проведения опытов представлена на рис. 5. Напряжение с выхода магнитофона

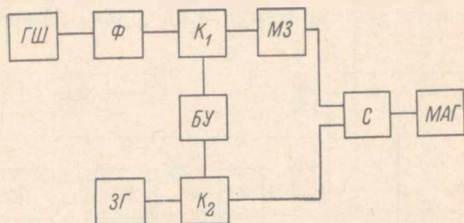


Рис. 3. Блок-схема установки для синтеза стимулов и записи их на магнитную пленку.

ГШ — генератор шума Г2-1; Ф — фильтр с полосой пропускания 1812—7240 гц; K_1, K_2 — электронные ключи; БУ — блок управления ключами; МЗ — магазин запусков МЗУ; ЗГ — звуковой генератор; С — суммирующий каскад; МАГ — магнитофон МЭБ-15.

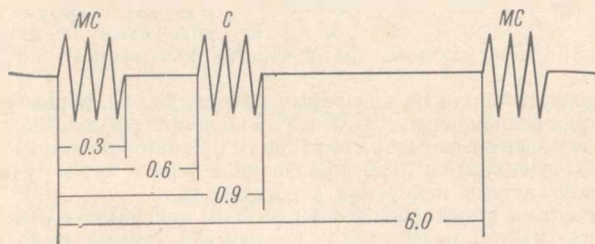


Рис. 4. Схема временной организации таблицы стимулов.

МС — маркирующий сигнал; С — стимул. Длительность временных интервалов в секундах.

подается одновременно в два канала: измерительный (на телефоны испытуемому через МЗУ) и в канал управления регистрирующей аппаратурой.

По окончании маркирующего сигнала блок управления (БУ) выдает импульс, переключающий двойной пиковый вольтметр (ПВ) в режим измерения и запоминания измеренных величин. Одновременно на реле Р подается импульс напряжения длительностью 600 мсек., так что в момент, когда на телефоны подается напряжение стимула, логарифмический усилитель (ЛУ) и через него вход 1 пикового вольтметра оказываются подключенными ко

входу телефона. Поданное на телефоны напряжение измеряется в логарифмическом масштабе, и максимальное его значение запоминается пиковым вольтметром.

По истечении 600 мсек. реле P обесточивается, логарифмический усилитель и пиковый вольтметр (вход II) подключаются к выходу измерительного усилителя акустического зонда (M , MU) и производится измерение и запоминание максимальной интенсивности сигнала, созданного испытуемым в ответ на данный стимул. Как только заканчивается измерение ответа испытуемого, вольтметр переключается

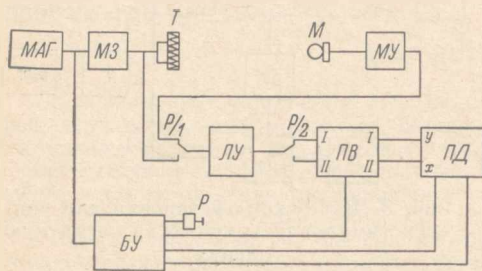


Рис. 5. Блок-схема установки для проведения опытов по имитации интенсивности шумных согласных.

M , MU — акустический зонд Мик-6с с усилителем УПУ-2; $ЛУ$ — логарифмический усилитель; $ПВ$ — двойной пиковый вольтметр ДПВ-3м (I — вход и выход первого канала, II — вход и выход второго канала); $ПД$ — потенциометр двухкоординатный ПДС-021 (y — вход первого канала, x — вход второго канала); $БУ$ — блоки автоматического управления системой измерения и регистрации; P — реле РЭС-9; $МАГ$ — магазин затухания МЗУ; T — телефон.

сится столько точек, сколько было ответов испытуемого на данный стимул (рис. 6).

После каждого опыта на диаграмму наносились калибровочные отметки по обоим координатным осям. Для этого на вход логарифмического усилителя подавали от звукового генератора тон 1000 гц при напряжении 100 мв, а затем, уменьшая его ступенями по 10 дб, производили запись путем ручного запуска автоматической системы измерения и регистрации.

На полученных описанным выше способом диаграммах измеряли интенсивности всех ответов испытуемого на каждый стимул и по полученным данным вычисляли для каждого испытуемого выборочные средние интенсивности ответов на каждый стимул и выборочные среднеквадратические отклонения ответов от среднего.

РЕЗУЛЬТАТЫ

Всего было проведено две серии экспериментов. В первой серии принимали участие четверо испытуемых (трое мужчин и одна женщина). Каждый испытуемый по несколько раз имитировал предъявленную ему исходную таблицу из 102 стимулов двумя способами: естественным (произносился шумный согласный [ʃ]) и путем продувания воздуха через дюзку постоянных размеров.

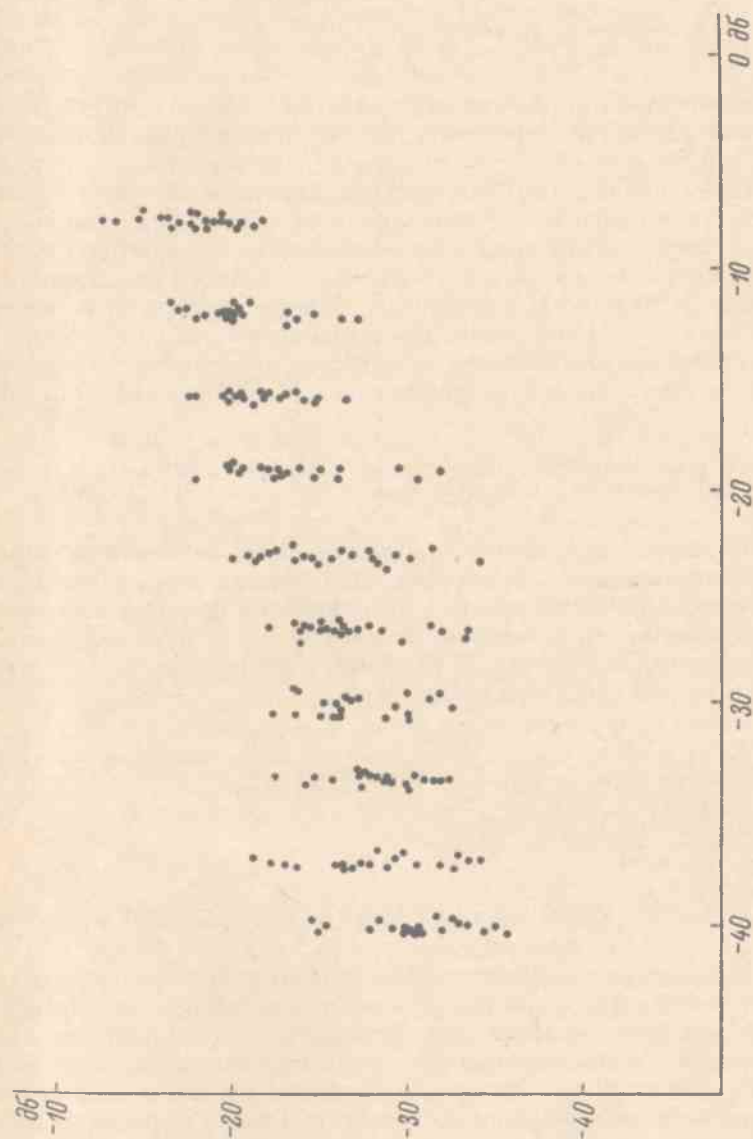


Рис. 6. Распределение ответов испытуемого (по оси ординат, $\delta\delta$) на стимулы разной интенсивности (по оси абсцисс, $\delta\delta$).

Для каждого испытуемого было получено следующее число ответов на каждый из предъявленных стимулов:

	М.	Г.	С.	Ж.	А.	В.	Н.	Ф.
В естественных условиях . . .	60	60	60	60	60	60	60	60
При имитации с помощью дюзы	80	60	30	60				

По измеренным интенсивностям ответов для каждого испытуемого были рассчитаны среднеквадратические отклонения ответов на каждый из стимулов при имитации в естественных и искусственных условиях. Так как среднеквадратические отклонения практически не зависят от интенсивности ответа (Малинникова, 1970), по этим данным вычислялось среднеквадратическое отклонение по всем ответам на все стимулы для данного испытуемого. Оказалось, что у всех испытуемых замена естественной щели в речевом тракте дюзой постоянных размеров приводит к уменьшению среднеквадратического отклонения ответов испытуемого. Ниже приводится среднеквадратическое отклонение ответов (в дБ):

	М.	Г.	С.	Ж.	А.	В.	Н.	Ф.
В естественных условиях . . .	3.06	2.73	3.32	2.49				
При имитации с помощью дюзы	1.88	2.60	2.62	2.16				

Чтобы оценить достоверность обнаруженных различий величин среднеквадратических отклонений, полученных при различных условиях эксперимента, для каждого значения среднеквадратического отклонения σ_1 , σ_2 были вычислены 5%-е границы для истинных среднеквадратических отклонений (в предположении нормального закона распределения):

	М. Г.	С. Ж.	А. В.	Н. Ф.
В естественных условиях	$2.91 < \sigma_1 < 3.24$	$2.53 < \sigma_1 < 2.89$	$3.15 < \sigma_1 < 3.52$	$2.36 < \sigma_1 < 2.64$
При имитации с помощью дюзы	$1.73 < \sigma_2 < 2.00$	$2.47 < \sigma_2 < 2.79$	$2.43 < \sigma_2 < 2.83$	$2.01 < \sigma_2 < 2.29$

Из приведенных данных следует, что во всех случаях, кроме одного (испытуемый С. Ж.), полученные различия величин среднеквадратических отклонений в разных условиях эксперимента являются достоверными. Следовательно, при замене естественной щели в речевом тракте во время имитации шумного согласного дюзой постоянных размеров происходит общее уменьшение разброса интенсивностей ответов испытуемых. Этот результат позволяет достаточно обоснованно утверждать, что интенсивность акустического сигнала, создаваемого испытуемым в ответ на услышанный стимул, непосредственно не регулируется. Чтобы оконча-

тельно убедиться в правильности этого вывода, сравним абсолютные величины интенсивностей ответов испытуемых на одинаковые стимулы при различных условиях эксперимента (рис. 7). Как видно из приведенных графиков, у всех испытуемых при имитации с помощью дюзы интенсивность ответов снижается, и это уменьшение интенсивности достоверно. Совершенно очевидно, что именно этого и следовало ожидать, если интенсивность создаваемого испытуемым акустического сигнала непосредственно не регу-

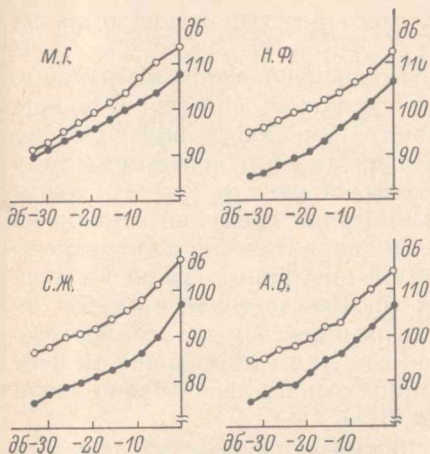


Рис. 7. Зависимость уровня интенсивности ответа (по оси ординат, дБ) от уровня интенсивности стимула (по оси абсцисс, дБ).

Кружки — результаты опытов при естественном произнесении; точки — результаты опытов с дюзой постоянных размеров.

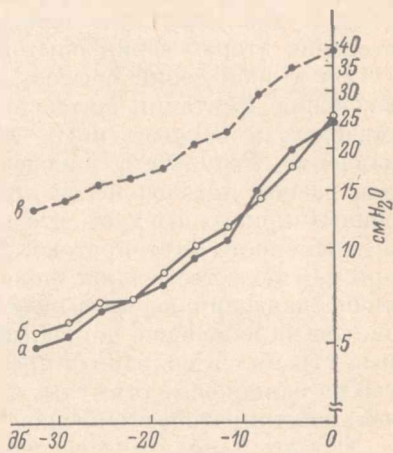


Рис. 8. Зависимость величины внутриротового давления (по оси ординат, см вод. ст.) от уровня интенсивности стимула (по оси абсцисс, дБ).

а — при имитации в естественных условиях; б — при имитации с помощью дюзы; в — давления в опытах с дюзой, необходимые для создания ответов, равных по интенсивности ответам в нормальных условиях.

лируется, поскольку к. п. д. дюзы как генератора шума значительно ниже, чем к. п. д. естественной щели (рис. 2).

Таким образом, можно считать установленным, что при имитации интенсивности синтетического шумного согласного испытуемые не регулируют непосредственно интенсивность создаваемого ими акустического сигнала, о чем свидетельствует значительное изменение параметров этого сигнала при искусственном изменении параметров щели в речевом тракте (замена естественной щели дюзой). По-видимому, испытуемый приводит в соответствие с интенсивностью услышанного стимула величину какого-то физиологического параметра. Тогда стоит попытаться выяснить, что это за параметр и как он может управляться.

Выше мы установили, что при имитации одинаковых стимулов интенсивность ответов испытуемого в опытах с дюзой значительно меньше, чем в нормальных условиях. Так как интенсивность шума, генерируемого в щели, определяется параметрами самой щели и величиной давления воздуха за щелью, а изменение параметров щели при замене естественной щели дюзой привело к изменению интенсивности генерируемого шума, можно было предположить, что при таком изменении параметров щели в речевом тракте осталось неизменным давление воздуха за щелью, т. е. внутриротовое давление. Для проверки этого предположения была поставлена вторая серия опытов.

Во второй серии опытов, проведенной на одном испытуемом, во время имитации интенсивности шумных согласных в естественных и искусственных условиях измерялось внутриротовое давление. Результаты измерений давления при имитации стимулов разной интенсивности приведены на рис. 8, а, б. Из приведенных кривых следует, что при имитации стимулов одинаковой интенсивности внутриротовое давление практически не изменяется при замене естественной щели дюзой постоянных размеров (различие давлений не превышает 17%). Вместе с тем, если бы, несмотря на изменение параметров щели в речевом тракте, испытуемый стремился сохранить прежние величины интенсивности ответов на одинаковые стимулы, ему пришлось бы значительно увеличить внутриротовое давление (рис. 8, в).

Остаются практически неизменными и среднеквадратические отклонения давлений в опытах с естественным произнесением и с дюзой (2.64 и 2.63 см H_2O соответственно).

Таким образом, полученные нами данные позволяют считать, что давление воздуха за местом образования щели (внутриротовое давление) может быть тем физиологическим параметром, который регулируется при имитации интенсивности шумных согласных.

Вопрос о том, является ли на самом деле внутриротовое давление этим регулируемым параметром, требует дополнительного исследования. В настоящее время об этом можно сказать лишь следующее. При произнесении изолированного глухого шумного согласного голосовые связки, как правило, находятся в разведенном состоянии, и суммарное аэродинамическое сопротивление гортани и нижних дыхательных путей мало по сравнению с сопротивлением щели, в которой генерируется шум. В этом случае при тех скоростях потока воздуха в речевом тракте, которые имеют место при произнесении согласных, внутриротовое давление практически равно внутрилегочному. Это значит, что, изменяя во время опытов по имитации интенсивности внутриротовое давление, мы на самом деле измеряли внутрилегочное. Тогда физиологическим параметром, управляемым при имитации интенсивности, будет внутрилегочное давление.

Но так как внутрилегочное давление создается за счет увеличения суммарного усилия всей дыхательной мускулатуры, то регулируемым параметром может оказаться именно усилие дыхательной мускулатуры, изменение которого в опытах по имитации интенсивности проявляется в виде изменения внутрилегочного и связанного с ним внутриротового давлений.

Л и т е р а т у р а

- Б е р н ш т е й н Н. А. 1947. О построении движений. М.
 М а л и н н и к о в а Т. Г. 1970. Связь между уровнями интенсивности стимула и ответа при имитации согласного. В сб.: Управление движениями Л.: 117—122.
 (F a n t G. 1960) Ф а н т Г. 1964. Акустическая теория речеобразования. М.
 М e y e r - E r p l e r W. 1953. Zum Erzeugungsmechanismus der Geräuschlaute. Ztschr. f. Phonetik, 7, 3—4: 196—212.

РАСПОЗНАВАНИЕ ОГРАНИЧЕННОГО НАБОРА УСТНЫХ КОМАНД

В. М. Величко, Н. Г. Загоруйко

Институт математики СО АН СССР, Новосибирск

Задача автоматического распознавания ограниченного набора устных команд привлекает внимание исследователей речевого сигнала главным образом с точки зрения получения практически приемлемых решений таких частных задач, как управление голосом автоматическими устройствами, устный ввод исходных данных и программ в ЭВМ, и т. д. (см., например, Высоцкий и др., 1968).

В данной статье описываются эксперименты по распознаванию словаря из 168 слов, список которых приводится ниже. В словарь вошли все термины языка «Алгол-60», некоторые термины входного языка α -транслятора и названия некоторых элементарных функций. Большинство названий букв русского и латинского алфавитов заменено собственными именами (как это принято в системах связи).

Для описания речевого сигнала в данном эксперименте была принята следующая система признаков. Акустический сигнал пропусклся через систему из 5 фильтров второго порядка. Центральные частоты фильтров были выбраны равными 112.5, 450, 900, 1800, 7200 гц (Курилов, Гаврилко, 1969). Ширина полосы пропускания каждого фильтра на уровне 6 дб равнялась центральной частоте фильтра, что соответствует добротности фильтров ~ 2.45 . Слово разбивалось на сегменты фиксированной длительности 14 мсек. В каждом сегменте подсчитывалась энергия сигнала в общей полосе E_0 и на выходе каждого из пяти фильтров E_i ($i=1, \dots, 5$). В качестве параметров, характеризующих сегмент, брались величины $\ln E_0$ и $\ln \frac{E_0}{E_i}$, т. е. сегмент описы-

вался точкой в шестимерном пространстве. Слово описывалось как последовательность сегментов. Частичное обоснование выбора логарифмических параметров состоит в том, что слух подчиняется известному в физиологии закону Вебера—Фехнера (Физический энциклопедический словарь, 1960), согласно которому минимально заметный прирост ΔI внешнего воздействия на органы чувств пропорционален величине воздействия I

$$\Delta I \sim \Delta \beta I,$$

откуда $\beta \sim \ln I$, т. е. сила восприятия пропорциональна логарифму внешнего воздействия.

СЛОВАРЬ «АЛГОЛ»

- | | | |
|---------------------|-----------------------------|-----------------------|
| 1. Один | 45. Набла | 89. Процедура |
| 2. Два | 46. Альфа | 90. Результат |
| 3. Три | 47. Бэта | 91. Значение |
| 4. Четыре | 48. Гамма | 92. Метка |
| 5. Пять | 49. Дельта | 93. Барабан |
| 6. Шесть | 50. Эпсилон | 94. Лента |
| 7. Семь | 51. Дзета | 95. Скаляр |
| 8. Восемь | 52. Эта | 96. Вектор |
| 9. Девять | 53. Тэта | 97. Матрица |
| 10. Ноль | 54. Каппа | 98. Строка |
| 11. Плюс | 55. Лямбда | 99. Функция |
| 12. Минус | 56. Ми | 100. Истина |
| 13. Разделить | 57. Нью | 101. Ложь |
| 14. Запятая | 58. Кси | 102. После |
| 15. Точка | 59. Пи | 103. Печать |
| 16. Пробел | 60. Ро | 104. Заменить |
| 17. Десять | 61. Сигма | 105. Число повторений |
| 18. Степень | 62. Тау | 106. Идентификатор |
| 19. Скобка | 63. Фи | 107. Конец отладки |
| 20. Открыть | 64. Хи | 108. Переключатель |
| 21. Закрывать | 65. Пси-штрих | 109. Присвоить |
| 22. Круглая | 66. Омега | 110. Дифференциал |
| 23. Умножить | 67. Аш | 111. Интеграл |
| 24. Равно | 68. Исправление и выделение | 112. Троеточие |
| 25. Точка с запятой | 69. Примечание | 113. Фигурная |
| 26. Квадратная | 70. Целый | 114. Синус |
| 27. Звездочка | 71. Вещественный | 115. Косинус |
| 28. Кавычки | 72. Комплексный | 116. Арксинус |
| 29. Не | 73. Логический | 117. Арктангенс |
| 30. Больше | 74. Собственный | 118. Логарифм |
| 31. Меньше | 75. Начало | 119. Экспонента |
| 32. Двоеточие | 76. Конец | 120. Корень |
| 33. Надчеркнуть | 77. Массив | 121. Русская |
| 34. Подчеркнуть | 78. Стоп | 122. Латинская |
| 35. Восклицательный | 79. Для | 123. Греческая |
| 36. Твердый | 80. Шаг | 124. Большая |
| 37. Градус | 81. До | 125. Малая |
| 38. Минута | 82. Цикл | 126. Аля |
| 39. Секунда | 83. Раз | 127. Борис |
| 40. Влево | 84. Пока | 128. Владимир |
| 41. Вправо | 85. Если | 129. Геннадий |
| 42. Вопросительный | 86. То | 130. Дмитрий |
| 43. Вниз | 87. Иначе | 131. Елена |
| 44. Плюс-минус | 88. На | 132. Женя |

133. Зинаида	145. Ульяна	157. Яков
134. Иван	146. Федор	158. Жи
135. И краткая	147. Харитон	159. Ку
136. Клим	148. Цецилия	160. Дубльвэ
137. Лидия	149. Чарли	161. Зэт
138. Михаил	150. Шура	162. Дизъюнкция
139. Николай	151. Щукарь	163. Конъюнкция
140. Ольга	152. Ы	164. Импликация
141. Павел	153. Мягкий	165. Целое деление
142. Роман	154. Знак	166. От
143. Сергей	155. Эдуард	167. Тожественно
144. Татьяна	156. Юрий	168. Отрицание

Для сравнения сегментов, принадлежащих различным реализациям, выбрана следующая мера сходства:

$$a = \frac{\alpha^2}{\rho^2 + \alpha^2}.$$

Здесь

$$\rho^2 = [\ln E_0^{(1)} - \ln E_0^{(2)}]^2 + \sum_{i=1}^5 \left[\ln \frac{E_0^{(1)}}{E_i^{(1)}} - \ln \frac{E_0^{(2)}}{E_i^{(2)}} \right]^2$$

является квадратом эвклидова расстояния между сегментами в шестимерном пространстве, α^2 принято равным 2 на основе предыдущих экспериментов (Загоруйко и др., 1968). Мера сходства может принимать значения от 1 для одинаковых сегментов до 0 для бесконечно удаленных.

Для сравнения слов использована идея, изложенная А. А. Пироговым (1968), которая заключается в следующем. Разные реализации одного и того же слова, как правило, имеют не совпадающую по времени структуру — одни фонемы удлиняются, другие — укорачиваются. Можно сравнивать слова посегментно — первый сегмент с первым, i -й с i -м, но более правильно сопоставлять друг с другом наиболее похожие сегменты (в смысле введенной меры сходства $[a]$), соответствующие одинаковым фонемам, и меру сходства между словами оценивать как сумму мер сходства между соответствующими сегментами с учетом длины слов. Возможная неоднозначность в определении меры сходства между словами при сравнении различных сегментов устраняется, если определить меру сходства между словами как максимально возможную сумму мер сходства между сегментами этих слов со следующими ограничениями: каждый сегмент участвует в сравнении не более одного раза, а порядок следования сегментов при сравнении соответствует порядку их следования в словах (чтобы не путать слова типа *то* и *от*). Для учета длины слов полученная сумма делится на длину максимального из двух сравниваемых слов. Описанная идея реализуется с помощью методов динамического программирования (см., например, Вентцель, 1964).

Приближенное вычисление меры сходства между словами можно осуществить следующим образом. Совмещаются начала и концы сравниваемых слов и суммируются меры сходства между сегментами, одинаково удаленными от начала слов. Для корректировки возможных ошибок при автоматическом определении границ слова эта процедура повторяется со словами, сдвинутыми друг от друга на некоторую величину. В нашем эксперименте сдвиг начала одного слова относительно другого составлял -8 , -4 , 0 , $+4$ и $+8$ сегментов, и из полученных мер сходства выбиралась максимальная.

Полностью алгоритм распознавания выглядел следующим образом. Предъявленное для распознавания слово сравнивалось приближенным методом с каждым из 168 слов-эталонов обучающей последовательности. Выбирались 16 слов обучающей последовательности, наиболее похожие на предъявленное слово. Затем методом динамического программирования проводилось точное определение меры сходства предъявленного слова с каждым из 16 эталонных слов и наиболее похожее слово-эталон выбиралось как результат распознавания. Комбинированная схема распознавания (Загоруйко, 1966) позволяет значительно сократить время счета при незначительном снижении надежности распознавания, так как время счета приближенной меры сходства значительно меньше, чем точной.

Статистические испытания проводились на следующем материале: один диктор (мужчина), 4 последовательности слов, из которых 3 включают по 168 слов, а 4-я — 166 слов (при вводе 4-й последовательности в ЭВМ были допущены ошибки в словах *kappa* и *tau*). Весь материал был записан на бьютовом магнитофоне «Астра-4» с микрофона типа МД-47 в тихой комнате без специальных условий для звукозаписи. Все 4 последовательности с магнитофона той же марки с помощью 9-разрядного аналого-цифрового преобразователя (частота квантования 20 кГц) были введены в ЭВМ БЭСМ-6, где производилось автоматическое членение на слова, выделение логарифмических параметров слов и запись слов на магнитную ленту БЭСМ-6.

Затем производилось распознавание, причем в качестве обучающих (эталонных) последовательностей использовались поочередно 1-я, 2-я и 3-я последовательности, а в качестве контрольных — все остальные. Результаты эксперимента приведены в таблице.

Число ошибок при автоматическом распознавании слов

Обучающая последовательность	Контрольная последовательность				Надежность, %
	1-я	2-я	3-я	4-я	
1-я		3	9	9	95.8
2-я	2		9	7	96.4
3-я	25	21		6	89.6

Общая надежность распознавания, усредненная по 1506 реализациям, составляет около 94%. Распознавание одного слова на ЭВМ БЭСМ-6 занимает в среднем около 8 сек. без учета времени работы внешних устройств (печать, магнитные ленты, магнитные барабаны) и около 10 сек. с учетом времени работы этих устройств.

Результаты экспериментов показывают, что с помощью данного метода можно получить практически приемлемую надежность распознавания 150—200 слов, если имеется возможность производить подстройку системы под диктора, состоящую в однократном произнесении всех слов словаря.

Дальнейшие исследования должны быть направлены главным образом на сокращение времени принятия решения.

Л и т е р а т у р а

- Вентцель Е. С. 1964. Элементы динамического программирования. М.
- Высоцкий Г. Я., В. Н. Трунин - Донской, Г. И. Цемель. 1968. Опознавание 40 слов с помощью цифровой вычислительной машины. Тр. VI Всесоюз. акуст. конфер., М.: ЗВ-3.
- Загоруйко Н. Г. 1966. Комбинированный метод принятия решений. В сб.: Вычислительные системы, Новосибирск, 22: 77—80.
- Загоруйко Н. Г., В. М. Величко, Г. Я. Волошин, В. Д. Гусев, В. Н. Елкина, И. В. Бахмутова, А. Г. Хайретдинова, Л. С. Юдина. 1968. Эксперименты по автоматическому распознаванию речевых сигналов. Доклад на IV школе-семинаре по распознаванию слуховых образов. Киев—Канев.
- Курйлов Б. М., Б. П. Гаврилко. 1969. Членение речевого потока на фонемы. В сб.: Вычислительные системы, вып. 35, Новосибирск (в печати).
- Пирогов А. А. 1968. К фонетической теории речи. Тр. VI Всесоюз. акуст. конфер., М.: ЗИВ-9.
- Физический энциклопедический словарь. 1960. 1: 242.
-

РЕФЕРАТЫ

УДК 612.821.8 : 414.41

О слуховом описании гласного. I. Признаки, различающие [i] и [e]. Мушников В. П., Чистович Л. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 5—10.

В статье описаны эксперименты, на основании которых делается вывод, что различение человеком гласных [i] и [e] производится по частоте первой форманты, граничное значение которой приблизительно равно 400 гц и не зависит от частоты второй форманты. Илл. — 5, библи. — 11 назв.

УДК 612.821.8 : 414.41

О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном. Мушников В. П., Чистович Л. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 11—19.

Результаты экспериментов, приведенных в статье, показывают, что для различения одноформантных гласных [u], [o] от двухформантных [i], [e], [ɔ], [ü] человек использует простой признак — наличие или отсутствие максимума в области частот выше 1000 гц. Илл. — 7, библи. — 11 назв.

УДК 612.821.8 : 414.41

Слуховое измерение первой форманты. Чистович Л. А., Шупляков В. С. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 19—28.

В работе описаны эксперименты, целью которых было установить, измеряет ли человек при классификации гласных величину форманты как резонансную частоту речевого тракта или использует результат измерения наиболее интенсивной гармоники в районе этого максимума. Результаты опытов показали, что человек пользуется приближенной оценкой величины форманты, ориентируясь на частоту наиболее выраженной спектральной составляющей. Это означает, что гипотеза измерений человеком формант путем анализа через синтез должна быть отвергнута. Илл. — 11, библи. — 18 назв.

УДК 612.858.001.572

Изучение реакции основной мембраны улитки на сложные акустические стимулы. Лабутин В. К., Павловский В. В. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 29—36.

На модели исследовалась реакция основной мембраны улитки на сложные акустические стимулы, четко различаемые человеком в психоакустических экспериментах. Установлено, что существующих представлений о частотной избирательности основной мембраны в сочетании с последующей нелинейной обработкой снимаемых с нее сигналов недостаточно для адекватного описания разрешающей способности слухового анализатора человека. Илл. — 7, табл. — 2, библи. — 13 назв.

Модель для описания результатов психоакустических экспериментов со стационарными сигналами. Тёмов В. Л. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 36—49.

Рассматривается модель, позволяющая вычислять пороги маскировки в психоакустических экспериментах. Сигнал описывается перечнем особенностей спектра типа срезов и максимумов («формант», выделяемых специальным механизмом. В работе модели проявляются некоторые эффекты, связываемые обычно с существованием критических полос слуховой системы. Модель, реализованная в виде программы для ЭВМ, позволяет работать со сложными сигналами, состоящими из большого количества тонов и полосовых шумов. Илл. — 6, библи. — 10 назв.

УДК 612.821.8 : 414.41

Восприятие коротких отрезков гласных звуков. Богданов Б. В. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 49—57.

Исследовалась разборчивость отрезков русских гласных, выделенных из стационарной части закрытых слогов. Длительность распознаваемых отрезков варьировала от 140 до 10 мсек. Обнаружена способность некоторых дикторов после значительной тренировки правильно идентифицировать гласные при их длительности порядка 15—20 мсек. Илл. — 1, библи. — 11 назв.

УДК 412.42 : 534.44

Акустический признак твердых и мягких согласных русского языка в связной речи (спектрографическое исследование). Шупляков В., Фант Г., Серна-Лейтао А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 57—66.

Цель работы состояла в проверке ряда гипотез относительно того, какие именно параметры акустического сигнала передают сведения о мягкости согласных. Для этого получены спектрограммы всех согласных русского языка, произнесенных 5 дикторами в составе слогов типа ГСГ. На спектрограммах измерены максимальные и минимальные значения частот второй и третьей формант на протяжении согласных или же в самом начале перехода к гласному. Обнаружено, что у мягких согласных частота второй форманты значительно выше, чем у твердых. Принимая за мягкие согласные только те, частота F_2 которых превышает некоторый пороговый уровень, различный для разных испытуемых, можно различать мягкие и твердые с надежностью не менее 96%. Пороговое значение частоты F_2 для каждого диктора может быть приблизительно вычислено, исходя из данных о среднем значении F_3 . Илл. — 9, табл. — 1, библи. — 8 назв.

УДК 612.821.8 : 534.32

Воспроизведение простых контуров изменения частоты основного тона звуков. Люблинская В. В. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 66—74.

Цель работы — выяснить, каким описанием пользуется человек при восприятии и воспроизведении изменяющейся во времени частоты основного тона звуков речи. В качестве методики исследования использована имитация человеком синтезированного гласного с линейным или скачкообразным изменением частоты основного тона. Сравнение исходных и ответных звуков позволило установить, что основными измеряемыми признаками таких контуров являются их начальные и конечные значения. Илл. — 5, табл. — 1, библи. — 6 назв.

Знак изменения частоты спектрального максимума как различительный признак места образования согласного. К о ж е в н и к о в В. А., Р е й т б л а т Л. Е., Ч и с т о в и ч Л. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 74—82.

Существующие экспериментальные данные позволяют думать, что признаком, различающим губные и язычные согласные является направление сдвига спектрального максимума на протяжении перехода. Это предположение проверено методом активного поиска границ на материале синтетических стимулов, имитирующих открытые слоги с начальным взрывным согласным. Показано, что граница, разделяющая губные и язычные согласные, совпадает с частотой спектрального максимума последующего гласного или же находится достаточно близко от него. Илл. — 6, табл. — 2, библи. — 14 назв.

УДК 612.85 : 534.32

Фонетическая интерпретация стимулов с резкими изменениями интенсивности. К у з ь м и н Ю. И., Л и с е н к о Д. М. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 83—92.

Исследовалось восприятие гласноподобных сигналов, на протяжении которых имелось резкое изменение интенсивности. Показано, что при определенной величине перепада интенсивности такие стимулы воспринимаются как открытые слоги, содержащие гласный и один из сонорных согласных. Эффект оказался одним и тем же при использовании стимулов с разными спектрами и разными абсолютными уровнями интенсивности. Илл. — 7, табл. — 2, библи. — 4 назв.

УДК 612.85 : 412.42 : 414.13

Восприятие длительности звонких согласных. Ж у к о в а М. Г. В сб.: Анализ речевых сигналов человека. 1971. Изд-во «Наука», Ленингр. отд., Л. 92—100.

Исследовалось восприятие и воспроизведение длительности звонкого участка русских начальных согласных [b], [d], [m] в слогах типа СГ. Использовались методы имитации и субъективного сравнения длительности. Полученные данные позволяют сделать вывод о том, что существует минимальная пороговая длительность звонкого согласного, ниже которой звонкий согласный воспринимается как соответствующий ему глухой. Эта длительность, равная в среднем 20 мсек., может рассматриваться как порог различения периодической структуры сигнала. Надпороговые длительности звонкого согласного измеряются начиная с величин, очень близких к пороговой, а средняя квадратическая ошибка членения не превышает 20 мсек. Илл. — 6, табл. — 1, библи. — 6 назв.

УДК 612.85-018.8 : 612.821.8.001.572

О функциональной модели механизма сегментации текущего потока сигналов. М о л ч а н о в А. П. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 100—109.

Высказывается предположение о существовании в слуховой системе механизма сегментации текущих сигналов. Рассматривается модель, описывающая отклики нейронов фазического типа и формируется функциональная схема механизма сегментации. С помощью предложенной модели удается разделить на фонемы фразу, состоящую из четырех слов. Илл. — 3, табл. — 1, библи. — 13 назв.

Восприятие ударения в простой звуковой последовательности. Федорова Н. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 110—122.

На материале восприятия простых звуковых последовательностей типа [aga] была проверена гипотеза об использовании человеком отношения звуковых энергий в качестве «меры» ударения. Полученные данные не позволяют принять эту гипотезу. При суждении о месте ударения человек учитывает и отношение длительностей гласных и отношение их интенсивностей; при этом отношение длительностей имеет значительно больший вес, чем отношение интенсивностей. Илл. — 7, библи. — 18 назв.

УДК 612.821.8 : 612.789

К вопросу о процедуре распознавания фонем. Чистович Л. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 123—136.

Предполагается, что переход от слухового описания речевого сигнала к фонемному символу осуществляется в два этапа: переход от описания сигнала в терминах акустических признаков к его описанию в терминах сходства с возможными фонемами и принятие решения о фонеме. Делается предположение о возможном виде функции сходства (зависимости степени сходства сигнала с той или иной фонемой от его слухового описания). Проведенное исследование показало возможность применения методов активного поиска границ и психологического шкалирования для изучения функций сходства. Илл. — 8, библи. — 13 назв.

УДК 612.821.8 : 414.41

О положении фонемной границы между синтетическими гласными [i]—[e], [ü]—[ö] и [u]—[o]. Слепокурова Н. А. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 136—138.

Проверялось предположение о том, что граница по первой форманте является общей для пар гласных [i]—[e], [ü]—[ö] и [u]—[o]. Данные, полученные методом активного поиска границ между этими гласными, не противоречат принятому предположению. Делается вывод о необходимости изучения закономерностей смещения фонемной границы в эксперименте. Илл. — 1, библи. — 7 назв.

УДК 612.821.8 : 15.073

Изменение положения фонемной границы во время эксперимента. Клевещев А. С. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 138—142.

Рассматриваются результаты экспериментов по активному поиску фонемной границы между гласными [i] и [e] при изменении частоты первой форманты. Показано, что результаты этих экспериментов не соответствуют принятому в литературе предположению о случайной флюктуации фонемной границы вокруг некоторого постоянного значения. Делается предположение о неслучайном изменении положения фонемной границы. Илл. — 4, библи. — 3 назв.

Контекстуальные влияния в экспериментах по определению фонемных границ. Кузьмин Ю. П., Лисенко Д. М. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 142—149.

Получены данные, свидетельствующие о том, что положение фонемной границы у каждого данного слушателя не является фиксированным, но флюктуирует в зависимости от ансамбля предъявляемых сигналов. Этот эффект наблюдается как при идентификации речеподобных стимулов, так и при идентификации естественно звучащих синтетических гласных. Илл. — 4, библи. — 11 назв.

Оценка голоса диктора по изолированным гласным звукам. Шейкин Р. Л. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 149—155.

Исследовалось распознавание мужских и женских голосов на материале изолированно произнесенных гласных. В работе использован метод шкалирования субъективных расстояний. Показано, что надежность распознавания голоса диктора существенно зависит от фонемного качества гласного. Особенно отчетливо эта зависимость обнаруживается при применении шепотных гласных. Из трех исследованных гласных наиболее надежное распознавание наблюдалось в случае гласного [i]. Илл. — 5, табл. — 1, библи. — 3 назв.

Шкалирование субъективных расстояний между синтетическими гласными. Голузина А. Г. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 155—160.

Методом шкалирования субъективных расстояний исследовалось, каким образом человек производит сравнение синтетических гласных при определении степени различия между ними. Результаты проведенных экспериментов позволяют заключить, что измерение субъективного расстояния между гласными производится после того, как произведена фонемная их классификация. Если гласные принадлежат к различным фонемным классам, то сравниваются классы, если же к одному классу, то сравниваются более подробные описания сигналов. Илл. — 5, библи. — 2 назв.

Об интонационном членении некоторых типов предложений русского языка. Кривнова О. Ф., Венцов А. В. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 161—172.

С помощью безынерционного измерителя периода колебаний голосовых связок экспериментально изучены интонационные кривые некоторых типов простых предложений русского языка. Установлено, что по виду полученных кривых предложение может быть разделено на определенное число интонационных единиц. Образующиеся при таком членении единицы соответствуют существующему в литературе определению синтагмы. Илл. — 5, библи. — 16 назв.

Исследование восприятия русских согласных методом серийного запоминания слогов. Г а л у н о в В. И., Ш н е й д е р В. Ю. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука». Ленингр. отд., Л. 172—177.

Рассматривается вопрос о способе записи согласных в кратковременной памяти. Анализ ошибок, допускаемых испытуемыми в экспериментах по серийному запоминанию слогов, позволяет думать, что в процессе эксперимента запоминаются не сами согласные фонемы, а их описание по некоторым субъективным признакам, вероятнее всего артикуляторным. При переходе от такого описания звука к решению о его принадлежности к той или иной фонеме человек, видимо, использует априорные вероятности встречаемости отдельных признаков. Табл. — 5, библи. — 3 назв.

УДК 153.8:414.42

Исследование субъективного представления группы русских согласных методом семантически противоположных пар. Г а л у н о в В. И. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 177—181.

Ставилась задача получить данные о структуре пространства субъективных признаков для группы русских согласных. Обосновывается применимость для этой цели метода семантически противоположных пар. Этим методом получены субъективные расстояния между исследованными согласными по ряду признаков. Сопоставить эти субъективные оси с какими-либо объективными признаками согласных можно было только для признака взрывные—щелевые. Илл. — 1, табл. — 1, библи. — 3 назв.

УДК 612.821.8:414.73:414.61

Восприятие фонем в слогах различных типов. Б о н д а р к о Л. В., К у к о л ь щ и к о в а Л. Е., П а в л о в а Л. П., С в е т о з а р о в а Н. Д., Ш т е р н А. С. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 182—192.

Анализируется распознавание открытых слогов, произнесенных в разных положениях относительно ударения. Слоги выделялись из слов при помощи сепаратора и предъявлялись слушателям для идентификации. Результаты показали, что лучше всего распознаются ударные и первые предупредительные слоги, хуже всего — заударные. Надежность распознавания безударных слогов заметно зависит от качества гласного. Обсуждается вопрос о характере редукции безударных слогов, содержащих разные гласные. Илл. — 4, табл. — 2, библи. — 9 назв.

УДК 612.782:534.79

Исследование механизма регулирования интенсивности шумных согласных. М а л и н н и к о в а Т. Г., В е н ц о в А. В. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 193—201.

Методом имитации исследовался механизм регулирования человеком интенсивности шумных согласных. Оказалось, что при создании звука требуемой интенсивности человек не использует акустическую обратную связь. Более вероятно, что при регулировании интенсивности шумных звуков задаются определенные величины физиологических параметров (внутрилегочное давление, активность дыхательной мускулатуры и т. д.). Илл. — 8, библи. — 4 назв.

Распознавание ограниченного набора устных команд. Величко В. М., Загоруйко Н. Г. В сб.: Анализ речевых сигналов человеком. 1971. Изд-во «Наука», Ленингр. отд., Л. 201—205.

Описываются эксперименты по распознаванию словаря, состоящего из 168 слов. Согласно принятой процедуре, каждое слово описывается последовательностью сегментов фиксированной длительности. В качестве параметров, характеризующих сегмент, используется общая энергия сигнала и его энергия в каждой из 5 полос. Аналогичным образом описываются создаваемые в ходе предварительного обучения эталоны. Выбор наиболее похожего эталона осуществляется в 2 этапа: сначала выбирается 16 наиболее близких слов, затем методом динамического программирования отбирается наиболее похожее слово. Процедура реализована на БЭСМ-6. Надежность распознавания, полученная в эксперименте, составила около 94%. Табл. — 1, библи. — 7 назв.

СОДЕРЖАНИЕ

	Стр.
Предисловие	3
I. Слуховое описание речеподобных сигналов (полезные психоакустические признаки)	
В. Н. Мушников, Л. А. Чистович. О слуховом описании гласного. I. Признаки, различающие [i] и [e]	5
В. Н. Мушников, Л. А. Чистович. О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном	11
Л. А. Чистович, В. С. Шупляков. Слуховое измерение первой форманты	19
В. К. Лабути́н, В. В. Павловский. Изучение реакции основной мембраны улитки на сложные акустические стимулы	29
В. Л. Те́мов. Модель для описания результатов психоакустических экспериментов со стационарными сигналами	36
Б. В. Богданов. Восприятие коротких отрезков гласных звуков	49
В. Шупляков, Г. Фант, А. Серпа-Лейтао. Акустический признак твердых и мягких согласных русского языка в связной речи (спектрографическое исследование)	57
В. В. Люблинская. Воспроизведение простых контуров изменения частоты основного тона звуков	66
В. А. Кожевников, Л. Е. Рейтблат, Л. А. Чистович. Знак изменения частоты спектрального максимума как различительный признак места образования согласного	74
Ю. И. Кузьмин, Д. М. Лисенко. Фонетическая интерпретация стимулов с резкими изменениями интенсивности	83
М. Г. Жукова. Восприятие длительности звонких согласных	92
А. П. Молчанов. О функциональной модели механизма сегментации текущего потока сигналов	100
Н. А. Федорова. Восприятие удара в простой звуковой последовательности	110
	213

II. Правила фонетической интерпретации и запоминаемое описание речевого элемента

Л. А. Чистович. К вопросу о процедуре распознавания фонем	123
Н. А. Слепокурова. О положении фонемной границы между синтетическими гласными [i]—[e], [ü]—[ö] и [u]—[o].	136
А. С. Клещев. Изменение положения фонемной границы во время эксперимента	138
Ю. И. Кузьмин, Д. М. Лисенко. Контекстуальные влияния в экспериментах по определению фонемных границ	142
Р. Л. Шейкин. Оценка голоса диктора по изолированным гласным звукам	149
А. Г. Голузина. Шкалирование субъективных расстояний между синтетическими гласными	155
О. Ф. Кривнова, А. В. Венцов. Об интонационном членении некоторых типов предложений русского языка	161
В. И. Галунов, В. Ю. Шнейдер. Исследование восприятия русских согласных методом серийного запоминания слогов	172
В. И. Галунов. Исследование субъективного представления группы русских согласных методом семантически противоположных пар	177
Л. В. Бондарко, Л. Е. Кукольщикова, Л. П. Павлова, Н. Д. Светозарова, А. С. Штерн. Восприятие фонем в слогах различных типов	182
Т. Г. Малинникова, А. В. Венцов. Исследование механизма регулирования интенсивности шумных согласных	193
В. М. Величко, Н. Г. Загоруйко. Распознавание ограниченного набора устных команд	201
Рефераты	206

АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ ЧЕЛОВЕКОМ

Проблемы физиологической акустики.

Том VII

Утверждено к печати

*Научным советом по физиологии органов чувств
Академии наук СССР*

Редактор издательства *И. М. Горбунова*

Художник *В. В. Грибакин*

Технический редактор *О. А. Мокеева*

Корректоры *Г. А. Александрова, А. А. Гинзбург*
и *О. И. Иващенко*

Сдано в набор 20/I 1970 г. Подписано к печати
1/III 1971 г. Формат бумаги $60 \times 90^{1/4}$. Бум. л. $6^{3/4}$.

Печ. л. $13^{1/2} = 13^{1/2}$ усл. печ. л. Уч.-изд. л. 14,36.

Изд. № 4313. Тип. зак. № 740. М-27070. Бумага № 2.

Тираж 2000. Цена 1 р. 55 к.

Ленинградское отделение издательства «Наука»
Ленинград, В-164, Менделеевская лин., д. 1

1-я тип. издательства «Наука».
Ленинград, В-34, 9 линия, д. 12

ИЗДАТЕЛЬСТВО «НАУКА»

*В магазинах конторы «Академкнига»
имеются в наличии книги:*

Винников Я. А., Титова А. К. Кортиев орган. Гистофизиология и гистохимия. 1961. 260 стр. Цена 1 р. 93 к.

Коштойац Х. С. Основы сравнительной физиологии. Изд. 2-е, доп. и перераб.

Т. 2. Сравнительная физиология нервной системы. 1957. 653 стр., 4 вкл. Цена 75 к.

Коштойац С. Х. Проблемы энзимохимии процессов возбуждения, торможения и эволюции функций нервной системы. (Баховские чтения. XVII). 1963. 31 стр. Цена 14 к.

Мчедlishvili Г. П. Функция сосудистых механизмов головного мозга и их роль в регулировании и в патологии мозгового кровообращения. 1968. 203 стр., 4 вкл. Цена 1 р. 23 к.

Орбели Л. А. Избранные труды. В пяти томах.

Т. 1. Вопросы эволюционной физиологии. 1961. 456 стр., 8 вкл. Цена 2 р. 85 к.

Т. 2. Адаптационно-трофическая функция нервной системы. 1962. 608 стр. Цена 3 р. 90 к.

Т. 3. Вопросы высшей нервной деятельности и ее развития. 1964. 480 стр. Цена 3 р. 08 к.

Т. 4. Вопросы общей физиологии и патофизиологии. 1966. 299 стр. Цена 1 р. 90 к.

Т. 5. Статьи и выступления. 1968. 320 стр. Цена 2 р. 23 к.

Первичные процессы в рецепторных элементах органов чувств. 1966. 196 стр. Цена 42 коп.

Физиологическая акустика. Библиографический указатель советской литературы. 1917—1950. 1960. 137 стр. Цена 47 к.

Заказы просим направлять по адресу:

Москва, В-463. Мичуринский проспект, дом № 12,
магазин «Книга — почтой».

Ленинград, П-110. Петрозаводская улица, дом № 7,
магазин «Книга — почтой».

ИСПРАВЛЕНИЯ И ОПЕЧАТКИ

Страница	Строка	Напечатано	Должно быть
21	6 сверху	изменялась	измерялась
34	Подпись к рис. 5, 2 сверху	A1.	A4.
34	Подпись к рис. 6, 2 сверху	A4.	A1.
76	19 снизу	{И}-образный	{U}-образный
78	Табл. 2, графа 2 справа, головка	»	«
112	18 снизу	{ага}.	{ага}.
191	5 сверху	предураные	предударные

Анализ речевых сигналов