

УДК 534.784

ВЛИЯНИЕ ВЫСОТЫ ОСНОВНОГО ТОНА НА ПОЛОЖЕНИЕ ФОНЕМНОЙ ГРАНИЦЫ МЕЖДУ ГЛАСНЫМИ

Н. А. Слепокурова

Исследована зависимость фонемной границы между гласными по частотному положению первого спектрального максимума (слуховому эквиваленту F_1) от высоты звука (слухового эквивалента F_0). В качестве стимулов использовались синтетические гласные (а) и (о) с фиксированной частотой F_2 и переменным максимумом в области F_1 , создаваемым с помощью 3, 4, 5 или 6-й гармоники F_0 . Показано, что с возрастанием F_0 граница между гласными по частоте первого спектрального максимума сдвигается вверх, но очень медленно: при удвоении F_0 сдвиг границы не превышает 5%.

Одной из трудностей автоматического распознавания гласных является большой разброс значений формантных частот у дикторов. Особенно большая разница существует между мужчинами, женщинами и детьми. Так, по данным Петерсона и Барни [1] для американско-английских гласных у женщин в среднем формантные частоты на 20% выше, чем у мужчин, а у детей восьмилетнего возраста средние формантные частоты на 20% выше, чем у женщин. Подобные же результаты получены для шведских [2] и японских [3] гласных. Во всех этих случаях, однако, имеется корреляция между положением формантных частот и частотой основного тона дикторов.

В автоматическом распознавании давно распространена идея об использовании частоты F_0 в качестве дополнительного параметра при распознавании речи. Интересно знать, что в этом случае происходит при восприятии, т. е. не являются ли у человека границы между гласными фонемами функциями от частоты F_0 . В литературе известны некоторые данные в пользу этого предположения. Прежде всего сюда относится работа Миллера [4].

В опытах Миллера использовался набор синтетических двухформантных гласных, полученных с помощью гармонического синтезатора, имеющего независимый контроль гармоник по амплитуде. Значения формант этих гласных образовывали матрицу в плоскости F_1F_2 , частота F_1 менялась шагами от 400 до 1000 гц частота F_2 — от 500 до 1000 гц. Гласные эти дважды предъявлялись испытуемым, от которых требовалось отождествить их с известными английскими фонемами. В первом случае расстояние между отдельными гармониками предъявляемых синтетических гласных, соответствующее частоте основного тона, составляло 144 гц, во втором случае с помощью устранения из спектра гласных каждой второй гармоники частота F_0 удваивалась, а спектральные огибающие оставались прежними. Результаты экспериментов по идентификации показали, что с удвоением частоты основного тона область, занимаемая каждой гласной в плоскости F_1F_2 , также сдвигается вверх. При этом сдвиг границы по первой форманте между звуками (U) и (A), проявившийся наиболее отчетливо, составляет

приблизительно 80 *гц*, что составляет 15% от первоначального значения границы по F_1 .

Зависимость между частотой F_0 и положением фонемных границ рассматривали также Фужисаки и Кавашима [3]. В их опытах с помощью метода идентификации определялись границы по F_1 между синтетическими японскими гласными (e) — (u) и (a) — (o) при меняющейся частоте F_0 . Оказалось, что с возрастанием частоты F_0 граничные значения F_1 для этих гласных также резко сдвигаются вверх: вычисления, проделанные с помощью приводимых в работе [3] графиков, показывают, что с удвоением частоты основного тона граница между гласными (a) и (o) по первой форманте сдвигается в среднем на 19%. Однако эти данные нельзя считать надежными, так как частота F_0 не являлась в эксперименте единственным переменным параметром: по определенному закону в зависимости от F_1 менялась также частота F_2 , которая могла по-своему влиять на положение границы по F_1 .

Весьма интересными также являются данные Слоусона [5]. В одном из его экспериментов испытуемым предъявлялись серии синтетических гласных, состоящие из 8 пар звуков. Первый звук в каждой паре был одинаковым для всей серии и соответствовал по спектральному составу одной из 6 английских гласных. Вторым звуком пары в каждой серии имел по сравнению с первым удвоенную частоту F_0 и измененную огибающую спектра, которая образовывалась путем умножения частот F_1 и F_2 первого звука на некоторый постоянный множитель. Всего было использовано 8 значений этого множителя (по числу пар в каждой серии) в диапазоне от 0,95 до 1,30. От испытуемых требовалось, не обращая внимания на высоту, оценить в баллах различие между гласными в каждой паре. После усреднения данных испытуемых для всех 6 серий выяснилось, что минимальные различия между гласными в паре наблюдались тогда, когда частоты F_1 и F_2 второго звука были в 1,1 раза больше чем соответственные частоты первого. Непосредственного изучения изменения границ между гласными под влиянием частоты F_0 в этих экспериментах, как видно, не производилось. Из работы [5] следует лишь то, что при удвоении частоты F_0 и одновременном увеличении частот первой и второй форманты на 10% фонемная принадлежность гласной не изменяется. Ясно, однако, что полученный эффект имел бы место и в том случае, если бы с возрастанием частоты F_0 происходил сдвиг фонемных границ.

Из всего изложенного следует, что, возможно, фонемная граница является функцией от частоты F_0 . В литературе нет никаких данных о виде этой функции, известно только, что она не убывает. Этому последнему условию удовлетворяет в математике широкий класс функций. В качестве одной из возможных аппроксимаций зависимости фонемной границы по какой-либо форманте от частоты F_0 удобно рассматривать степенную функцию. В этом случае

$$F_{гр} = KF_0^a,$$

где k — некоторая константа, равная $F_{гр}$ в случае, если $a = 0$. Существенным для нас здесь является показатель a , характеризующий степень возрастания границы в зависимости от F_0 ; очевидно, что $0 \leq a \leq 1$. Несложный расчет показывает, что при таком характере зависимости удвоению частоты основного тона соответствует сдвиг границы на $[(2^a - 1) \cdot 100]\%$. Действительно, пусть $F_{гр}' = k(F_0')^a$ и $F_{гр}'' = k(F_0'')^a$. Тогда $(F_{гр}'' - F_{гр}') / F_{гр}' = [k(F_0'')^a - k(F_0')^a] / k(F_0')^a = [(F_0'')^a - (F_0')^a] / (F_0')^a$. Но по условию $F_0'' = 2F_0'$, откуда $(F_{гр}'' - F_{гр}') / F_{гр}' = k(F_0')^a(2^a - 1) / k(F_0')^a = 2^a - 1$. Поэтому, например, зная, у Миллера с удвоением частоты F_0 граница по F_1 сдвигается на 15% и предположив, что найденная им зависимость $F_{гр}$ от F_0 , описывается введенной выше функцией, можно найти показатель a . В этом случае

$$F_{гр} \approx kF_0^{0.2}.$$

Результаты, полученные Миллером и Фужисаки и Кавашима, однако, имеют сложную природу. Их авторы исходили из предположения, что человек измеряет частоты F_1 и F_2 . Однако в последнее время получены [6, 7], показывающие, что при восприятии гласных частота F_1 непосредственно не измеряется, а ее эквивалентом для человека является частота наиболее громкой гармоники в той части спектра, которая соответствует первой форманте. Этот параметр в отличие от частоты F_1 уже не является независимым от F_0 ; частота гармоники есть функция от F_0 и от F_1 . Не исключена возможность, что сдвиг границ, наблюдавшийся в работах [3 и 4], происходил именно вследствие того, что измерялась не форманта, а соответствующая гармоника. Поэтому результаты работ [3 и 4] нужно рассматривать лишь как ориентировочные.

В свете данных работ [6 и 7] поставленный в начале статьи вопрос о зависимости фонемных границ от частоты F_0 можно переформулировать следующим образом: является ли граница между гласными по частотному положению первого спектрального максимума (слухового эквивалента F_1) функцией от высоты звука (слухового эквивалента F_0).

Для ответа на этот вопрос нужно было исследовать такие сигналы, которые давали бы возможность непрерывного изменения высоты и связанного с ним изменения частотного положения спектрального максимума. Очень полезными оказались данные работ [6 и 7], из которых следует, что замена всего спектра в области первой форманты единственной гармоникой не влияет на фонемное качество гласных. Оказалось возможным использовать при экспериментах синтетические гласные, у которых первая форманта заменялась спектральным максимумом, образованным той или иной гармоникой основного тона. Изменяя при фиксированной гармонике частоту F_0 , можно было тем самым изменять положение спектрального максимума и искать по нему границу между гласными.

В спектре гласных, использованных при опытах, было только два максимума. Первый образовывался указанным выше способом, в качестве второго выступала фиксированная вторая форманта, что позволяло обойтись без непосредственного присутствия в спектре частоты F_0 : согласно данным психоакустики частоты F_0 (высота сигнала) должна была восстанавливаться вследствие наличия периодичности при обработке временной структуры сигнала в области второй форманты.

Выбор гласных фонем, граница между которыми являлась объектом исследования, определялся следующими требованиями:

1) поскольку значение второй форманты должно было быть фиксированным, т. е. общим для обеих гласных, необходимо было использовать гласные с близкими значениями этого параметра;

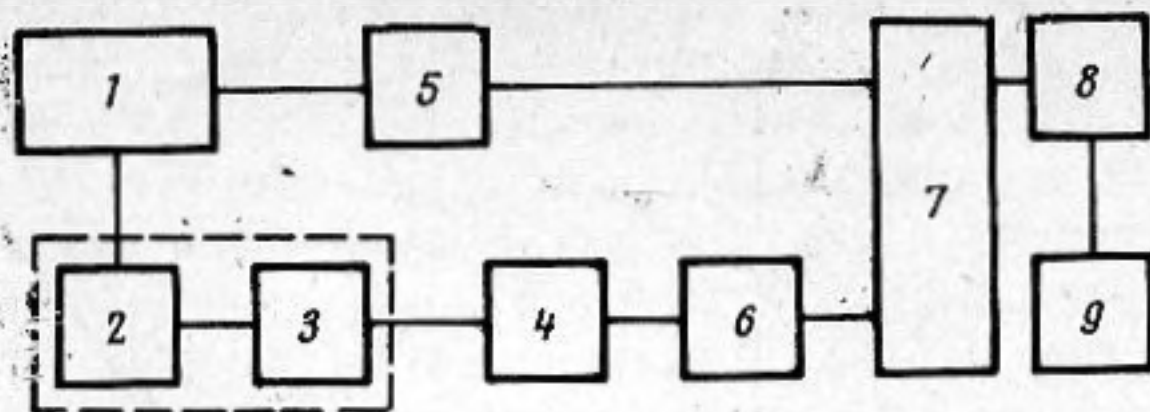
2) нужно было, чтобы предполагаемая граница по F_1 между этими гласными лежала достаточно высоко, т. е. на такой частоте, которая дала бы возможность использовать для получения спектрального максимума не менее 3—4 гармоник основного тона;

3) чтобы избежать возможного влияния частоты F_2 на положение границы, следовало использовать такие гласные, у которых частоты F_1 и F_2 достаточно далеко отстоят друг от друга.

Первые два требования являлись решающими. Им хорошо удовлетворяют гласные (а) и (о), которые и были выбраны для исследования. Вторая форманта, общая для обеих гласных, фиксировалась на частоте 1195 гц. Выбор столь высокой частоты F_2 объясняется соображениями, изложенными в последнем требовании, и оказался возможным лишь потому, что при данном значении второй форманты гласные (а) и (о) сохраняют свое фонемное качество.

Схема установки, с помощью которой синтезировались гласные, показана на фиг. 1. Источником гармоник служил генератор сложного напряжения ГСН-1 1, уровень которого регулировался с помощью магазина затуханий МЗУ 5. Источником второй форманты являлся формантный

синтезатор (обведен пунктиром); частота источника основного тона 2 управлялась с помощью первой гармоники от ГСН-1. Сигнал с выхода формантного фильтра 3 поступал через фильтр верхних частот 4 и магазин затуханий 6 на общий сумматор 7. Длительность сигнала 300 мсек и частота повторения 1 гц задавались с помощью электронного ключа 8 и схемы управления 9. Сигналы, получаемые с помощью этой установки, использовались в нескольких экспериментах.



Фиг. 1

Первый эксперимент был ориентировочным и служил для определения области, в которой могла находиться граница между гласными (а) и (о) и соответственно области изменения частот для 3, 4, 5 и 6 гармоник. Опыт проводился по методу активного поиска границы, заключавшемся в следующем. Сигнал с установки поступал на электродинамический головной

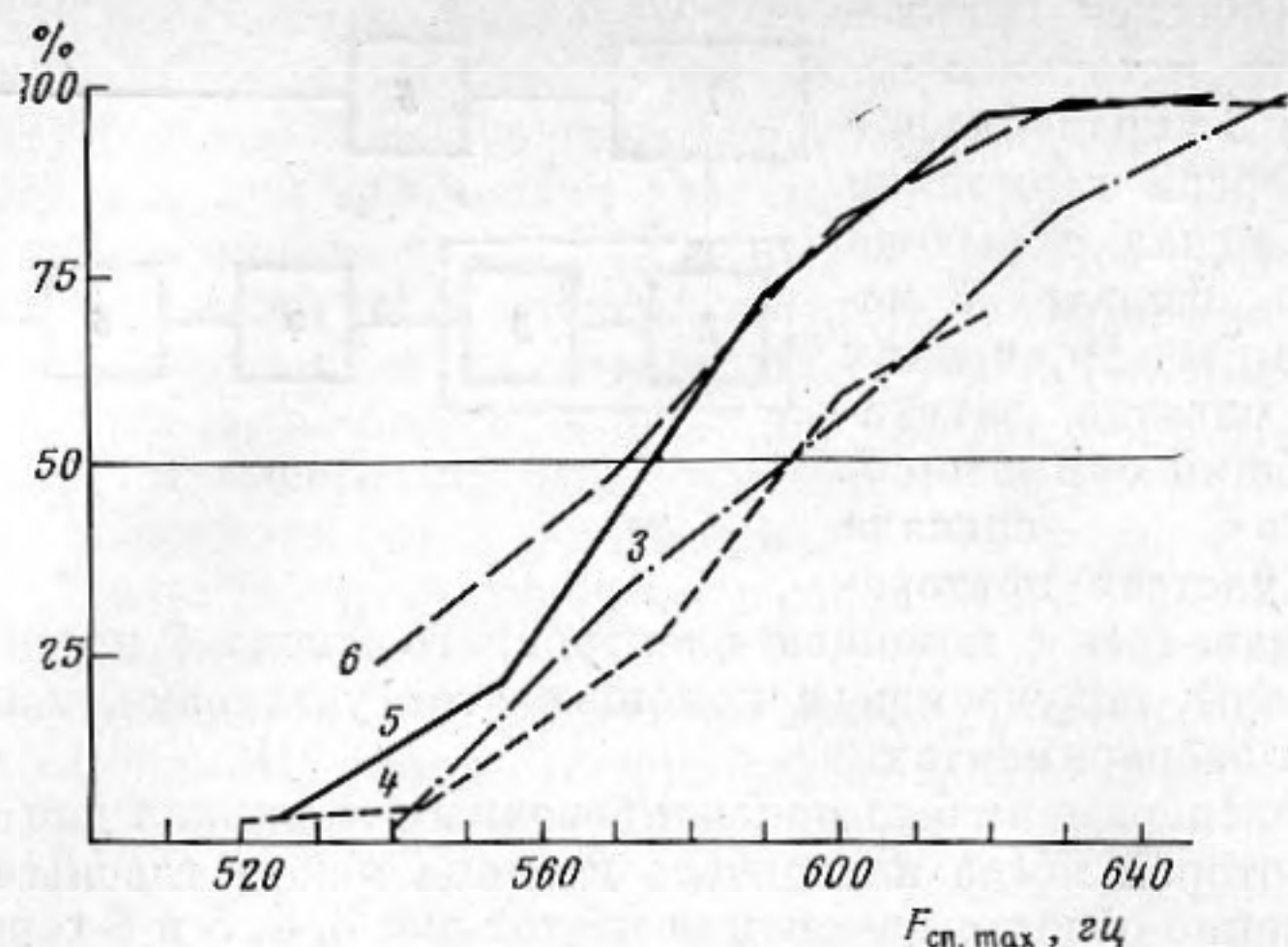
Таблица 1

F_0 исх.	135	125	150	120	105	170	115	90	200	140	190	102	210	105	180	145	93	130	94	110
F_0 имит.	126	121	155	127	107	170	117	104	188	141	184	100	202	102	160	145	99	132	101	160

телефон ТД-6 испытуемого. Экспериментатор включал одну из указанных выше гармоник, а испытуемый, плавно изменяя частоту основного тона, искал то ее значение, при котором получался звук, соответствующий, по мнению испытуемого, переходу от (а) и (о). В эксперименте участвовало трое испытуемых, каждый из которых произвел по 30—40 установок границы для каждой из 4-х гармоник. Результаты эксперимента показали, что положение границы между (а) и (о) по первому спектральному максимуму варьирует в диапазоне 500—650 гц, а значения частоты F_0 изменяются от 90 гц для 6-й гармоники до 210 гц для 3-й.

Второй эксперимент ставился для проверки того, что высота предъявляемых звуков действительно соответствует частоте F_0 , отсутствующей в их спектре, и проводился с помощью имитации. На магнитную ленту было записано 20 сигналов, лежащих в диапазоне границы по F_1 . Испытуемый прослушивал сигналы через телефоны и имитировал их по высоте. Основной тон имитированного сигнала снимался с помощью ларингофона и измерялся счетным частотомером ЧЗ-ЗА. В табл. 1 представлены частоты основного тона исходных (верхняя строчка) и имитированных (нижняя строчка) сигналов. Каждое значение частоты F_0 в нижней строчке является средним по 10 измерениям. Видно, что между значениями частоты F_0 в верхней и нижней строчках таблицы имеется хорошее соответствие. Небольшие отклонения значений F_0 имитированных сигналов от F_0 исходных сигналов могут быть отнесены к ошибкам воспроизведения.

Третий и четвертый эксперименты были основными и ставились непосредственно для поиска зависимости между фонемной границей по частоте первого спектрального максимума и высотой звука. Третий эксперимент проводился по методу идентификации. Сигналы записывались на магнитную ленту и затем через телефоны предъявлялись испытуемым. Последние должны были принять решение о том, каким гласным — (а) или (о) — является данный сигнал и соответственно повернуть ключ вправо или влево. Реакции испытуемых записывались на многоканальном самописце. Всего испытуемым предъявлялось 22 различных сигнала с одинаковым значением F_2 и разными значениями частоты первого спектрального



Фиг. 2

максимума ($F_{\text{сп.мах}}$, образованными с помощью 3, 4, 5 и 6-й гармоник основного тона. Области изменения F_0 и $F_{\text{сп.мах}}$ соответствовали тем, которые были найдены в первом эксперименте. Частоты F_0 и $F_{\text{сп.мах}}$ использованных гласных указаны в табл. 2. В экспериментах приняло участие 7 испытуемых, каждый из 22 сигналов прослушивался 35—40 раз.

Таблица 2

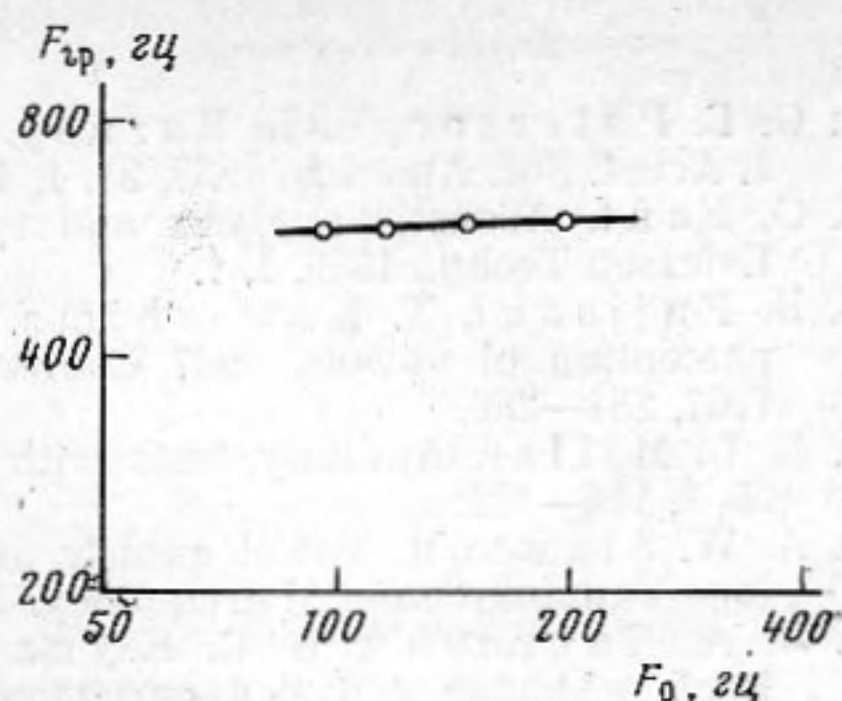
	6 гармоника					5 гармоника					4 гармоника					3 гармоника						
F_0	90	95	100	105	110	105	111	118	124	130	130	136	144	150	155	170	180	190	200	210	220	230
$F_{\text{сп.мах}}$	540	570	600	630	660	525	555	590	620	650	520	544	576	600	620	510	540	570	600	630	660	690

На фиг. 2 показаны результаты эксперимента по идентификации. Каждая точка ломаной линии, отвечающих 3, 4, 5 и 6-й гармоникам, является средним для 7 испытуемых и показывает процент гласных, принятых при данном значении $F_{\text{сп.мах}}$ за (а). Границей между (а) и (о) можно считать то значение $F_{\text{сп.мах}}$, которое соответствует 50% распознавания (а). Из фиг. 2 видно, что граничные значения частоты первого спектрального максимума ($F_{\text{гр}}$) для всех четырех гармоник расположены весьма близко друг к другу, диапазон разброса составляет приблизительно 20 гц. На фиг. 3 показана зависимость $F_{\text{гр}}$ для гласных (а) и (о) от частоты основного тона (частота по обеим осям отложена в логарифмическом масштабе). Кружочками показаны средние значения частоты $F_{\text{гр}}$ для четырех значений частоты F_0 , вычисленных по фиг. 2, сплошная линия является приближительной аппроксимацией зависимости $F_{\text{гр}}$ от F_0 . Из графика видно, что с возрастанием частоты F_0 частота $F_{\text{гр}}$ увеличивается, но очень мало. Точный расчет показывает, что наблюдавшийся в эксперименте по идентификации сдвиг границы по $F_{\text{сп.мах}}$ при увеличении частоты F_0 на октаву не превышает 5%.

Суммарные данные, полученные в опытах по идентификации, было решено проверить на отдельных испытуемых, используя метод активного поиска границы (четвертый эксперимент). В опытах участвовало 8 испытуемых, каждый из которых произвел по 20 установок границы. Испытуемый определял подряд два значения $F_{\text{гр}}$ для 4-й и два — для 6-й гармоники,

которые затем усреднялись соответственно для каждой гармоник. Затем на основании полученных значений F_0 и $F_{гр}$ вычислялся параметр a степенной кривой, аппроксимирующей полученную зависимость. Это делалось следующим образом: $4F_0'' = k(G_0'')^a$; $6F_0' = k(F_0')^a$; $4F_0'' / 6F_0' = (F_0'')^a / (F_0')^a$; $\lg 2F_0'' - \lg 3F_0' = a(\lg F_0'' - \lg F_0')$; $a = (\lg 2 - \lg 3 + \lg F_0'' - \lg F_0') / (\lg F_0'' - \lg F_0') = -0,175 / (\lg F_0'' - \lg F_0') + 1$.

В табл. 3 приводятся показатели a , найденные для 8 испытуемых по каждой отдельной паре 6-я — 4-я гармоника, а также их средние значения по пяти парам. Из таблицы видно, что средний показатель a очень резко отличается от остальных только для испытуемых под номерами 6 и 8, но их отклонения имеют разные знаки и, следовательно, компенсируются (знак минус перед a показывает, что при данном нахождении границы по 4 и 6 гармонике частота $F_{гр}$ оказалась больше для 6 гармоник; такие измерения границы противоречат общей тенденции возрастания $F_{гр}$ с возрастанием F_0 , но, очевидно, для всех испытуемых, кроме восьмого, являются случайными). Среднее по восьми испытуемым значение a составляет 0,04. Если отбросить данные испытуемых 6 и 8, то среднее a по шести остальным испытуемым также равно 0,04. Это означает, что при увеличении



Фиг. 3

Таблица 3

№ испытуемых

	1	2	3	4	5	6	7	8
a_1	-0,03	0,04	0,01	0,06	-0,07	0,18	-0,10	-0,12
a_2	-0,07	0,07	0,10	0,02	0,03	0,20	0,15	0,17
a_3	0,09	0,02	0,04	-0,11	0,05	0,11	-0,07	-0,11
a_4	0,14	0,11	-0,01	0,16	0	0,36	0,05	-0,15
a_5	-0,01	0,15	0,17	0,04	0,05	0,20	0,19	-0,50
a , средн.	0,024	0,078	0,062	0,034	0,012	0,210	0,044	-0,142

частоты F_0 на октаву граница между гласными (a) и (o) по $F_{сп\ max}$, найденная в экспериментах по методу активного поиска, возрастает в среднем на 3%. Эти данные хорошо согласуются с теми, которые были получены в эксперименте по идентификации.

Выводы

1. Можно считать, что действительно существует зависимость между фонемной границей гласных по положению первого спектрального максимума (слухового эквивалента F_1) и высотой (слуховым эквивалентом F_0). Однако эта зависимость является очень слабой, возрастание $F_{гр}$ на 3—5% при удвоении частоты F_0 , полученное в наших экспериментах, сравнимо с величиной дифференциального порога по частоте F_1 и F_2 , которая, по данным Фланагана [8], составляет 3—5%.

2. Не исключена возможность того, что небольшое увеличение частоты $F_{гр}$, которое наблюдалось в наших экспериментах, происходило вследствие незначительных изменений второй форманты, возникавших при изменении частоты основного тона. Эти изменения F_2 в данной работе во внимание не принимались.

ЛИТЕРАТУРА

1. G. E. Peterson, H. L. Barney. Control methods used in a study of the vowels. J. Acoust. Soc. America, 1952, 24, 1, 175—184.
2. G. Fant. Acoustic analysis and synthesis of speech with application to Swedish. Ericsson Techn., 1959, 1, 1.
3. H. Fujisaki, T. Kawashima. The roles of pitch and higher formants in the perception of vowels. 1967 Conference on Speech Communication and processing. 1967, 251—256.
4. R. L. Miller. Auditory tests with synthetic vowels. J. Acoust. Soc. America, 1952, 25, 1, 114—121.
5. A. W. Slawson. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. J. Acoust. Soc. America, 1968, 43, 1, 87—101.
6. Л. А. Чистович, В. С. Шупляков. Слуховое измерение первой форманты. В сб. «Анализ речевых сигналов человеком», Л., «Наука», 19—28.
7. В. Н. Мушников, Л. А. Чистович. Метод экспериментального исследования роли громкостей составляющих гласного в его распознавании. Акус., 1971, 17, 3, 405—411.
8. J. L. Flanagan. A difference limen for formant frequency. Journ. Acoust. Soc. America, 1955, 27, 3, 613—617.

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

Поступила в редакцию
3 марта 1971 г.