АКУСТИЧЕСКИЙ ЖУРНАЛ

Том VIII 1962 Вып. 1

ВРЕМЕННЫЕ ЗАДЕРЖКИ БУКВЕННОЙ ЗАПИСИ ГЛАСНЫХ ЧЕЛОВЕКОМ — НАБЛЮДАТЕЛЕМ

Р. О. Алекин, Ю. А. Клаас, Л. А. Чистович

Приводятся данные, показывающие, что временные задержки при буквенной записи гласных в среднем значительно превышают задержки повторения тех же гласных. Демонстрируется периодичность (с периодом около 100 *мсек*) в распределении времен задержек записи, анализируется зависимость характера распределения от качества гласной. При обсуждении данных используется модель, основанная на гипотезе о периодическом считывании информации в центральной нервной системе.

Постановка задачи автоматического распознавания звуков речи вызвала значительный интерес к особенностям процесса распознавания речи человеком, в частности к временным характеристикам этого процесса. Рядом авторов [1, 2, 3] было показано, что временные задержки повторения звуков речи очень малы, составляют всего 150—200 *мсек*. Однако дальнейшие исследования [4, 5] позволили прийти к выводу, что при повторении реализуется только первый этап процесса распознавания — определение характеристик звука по признакам, полезным в системе родного языка повторяющего *. Второй этап распознавания (выбор фонемы на основании найденных характеристик) при повторении не реализуется: найденные характеристики непосредственно используются для управления речевым аппаратом. Естественно поэтому, что компрессия речи, осуществляемая при повторении, оказывается значительно меньше той, которая может быть достигнута при полном фонемном распознавании. Полное фонемное распознавание неизбежно происходит в случае буквенной записи изолированных звуков. Представляло интерес выяснить, превышают ли временные задержки при записи букв временные задержки повторения.

Для магнитной записи и воспроизведения таблиц гласных использовался магнитофон МЭЗ-15. Испытуемый должен был как можно быстрее записывать буквой слышимый им гласный. Перемещения карандаша по бумаге при записи букв воспроизводились в виде изменений двух электрических сигналов, один из которых был пропорционален смещению карандаша по вертикальной, а другой — по горизонтальной оси листа бумаги. Оба этих сигнала, так же как и огибающая гласного звука, регистрировались параллельно на ленте чернильного осциллографа. В опытах участвовали двое испытуемых, каждый из которых записал по 900 естественных и по 1260 укороченных гласных.

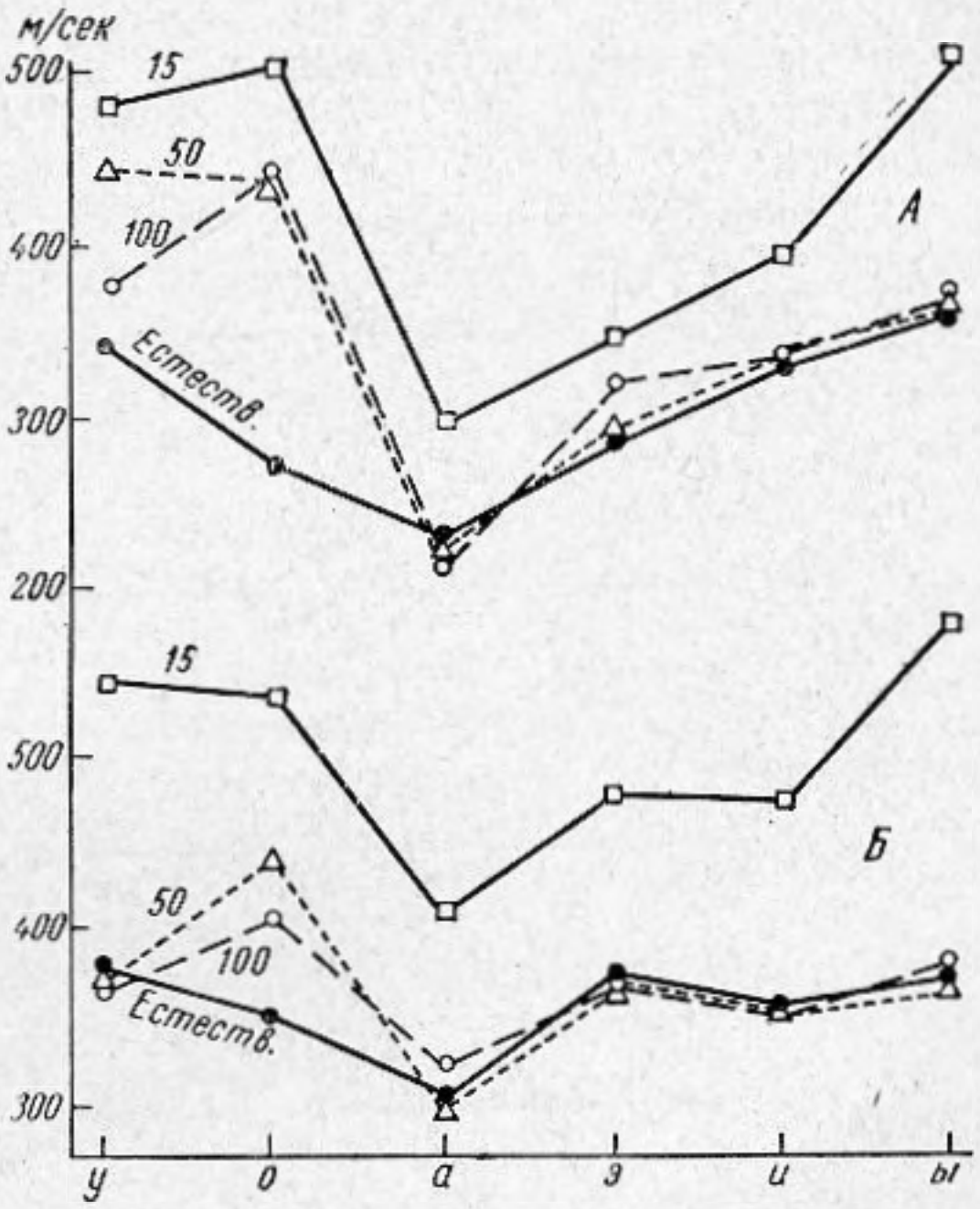
Данные, полученные в опытах с естественными гласными, показали прежде всего, что средние временные задержки буквенной записи весьма велики, составляя у испытуемого *А* — 315 *мсек*, у испытуемого *Б* — 366 *мсек*. Средние временные задержки имитации тех же звуков у тех же испытуемых составляют всего 163 *мсек* (*А*) и 148 *мсек* (*Б*), т. е. являются значительно меньшими. Так как увеличение времени задержки по сравнению с задержкой при имитации было получено также в тех условиях, когда человек должен был не писать букву, но называть фонему, с которой

---

* Логика процесса распознавания, содержание понятия «признак» и так далее подробно рассматриваются в работе Бонгарда [7].

совпадает слышимый звук [4], можно утверждать, что это увеличение задержки связано с процессом выбора фонемы (буквы), отсутствующем при имитации.

Ряд данных экспериментальной психологии (см. [7]) показывает, что, производя выбор, человек действует как одноканальная система (два выбора не могут осуществляться одновременно). Отсюда следует, что для того чтобы каждый звук речи в речевом потоке мог быть опознан как соответствующая фонема (буква), звуки должны следовать друг за другом не чаще чем через 300—400 *мсек* (время задержки, связанное с одним решением). Так как на самом деле средняя длительность звука в связной речи является значительно меньшей, составляет всего 100—200 *мсек*, приходится заключить, что человек не производит фонемного распознавания каждого отдельного звука в естественной речи. Можно думать, что при нормальном восприятии речи процесс выбора осуществляется на уровне более крупных единиц—последовательностей звуков. Сведения о звуках запоминаются до момента выбора в форме набора характеристик этих звуков по их полезным признакам.

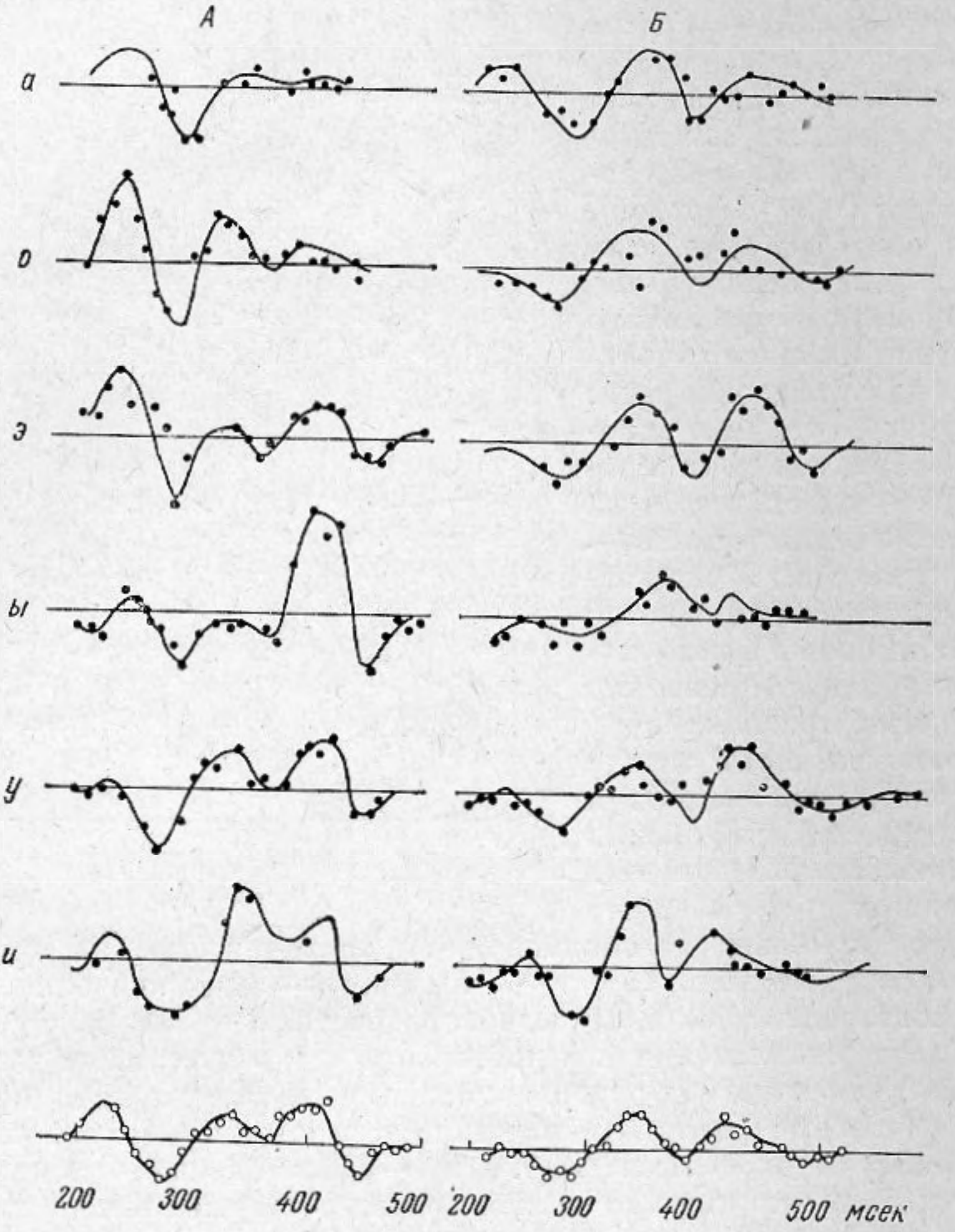


Фиг. 1

Установив разность между средней задержкой записи и средней задержкой повторения (Δt), мы попытались проверить простую гипотезу о том, что эта разность отражает постоянное время, затрачиваемое на выбор фонемы и что, следовательно, $t_з = t_п + \Delta t$, где $t_з$ — задержка записи, $t_п$ — задержка повторения. Были вычислены разности задержек между самыми короткими задержками записи и повторения и между самыми длинными задержками. Для коротких задержек средняя (из 20) разность оказалась равной 55 *мсек* (*А*) и 74 *мсек* (*Б*), для длинных задержек разность оказалась равной 210 *мсек* (*А*) и 376 *мсек* (*Б*), т. е. в 4—5 раз большей. Эти цифры заставляют считать, что нижний предел времени, затрачиваемого на выбор фонемы, не превышает 55 *мсек*, но что это время может колебаться в очень широких пределах.

Для выяснения факторов, влияющих на время выбора, мы обратились прежде всего к сравнению средних задержек записи разных естественных гласных. Полученные данные приведены на фиг. 1, где на оси ординат отложены средние задержки в миллисекундах; параметр кривых характеризует длительность звука. Из графика следует, что у обоих испытуемых наименьшие задержки наблюдались при гласной *а*, наибольшие — при гласных *у* и *ы*. Казалось возможным предположить, что эта разница в задержках связана с различным распределением информации на протяжении гласного и, соответственно, с различным, для разных гласных, временем накопления информации до принятия решения. Для того чтобы проверить это предположение, были проведены опыты с укороченными гласными. Укорочение звуков достигалось с помощью реле, управляемого началом звука и выключающего звук через заданный интервал времени (15, 50 и 100 *мсек*). Показателем искажений звука вследствие укорочения служило удлинение времени задержки буквенной записи.

Согласно высказанной выше гипотезе следовало ожидать, что чем длиннее задержка записи естественного гласного, тем больше его полезная длительность и, следовательно, тем при меньшем его укорочении будут наступать искажения. Полученные данные (фиг. 1) показали, что такого параллелизма не наблюдается. При укорочении до 100 *мсек* отчетливо пострадал только один звук *о*, хотя этот звук был связан с наименьшими, после *а*, временными задержками. На *а*, *э*, *и*, *ы* не оказало никакого влияния и

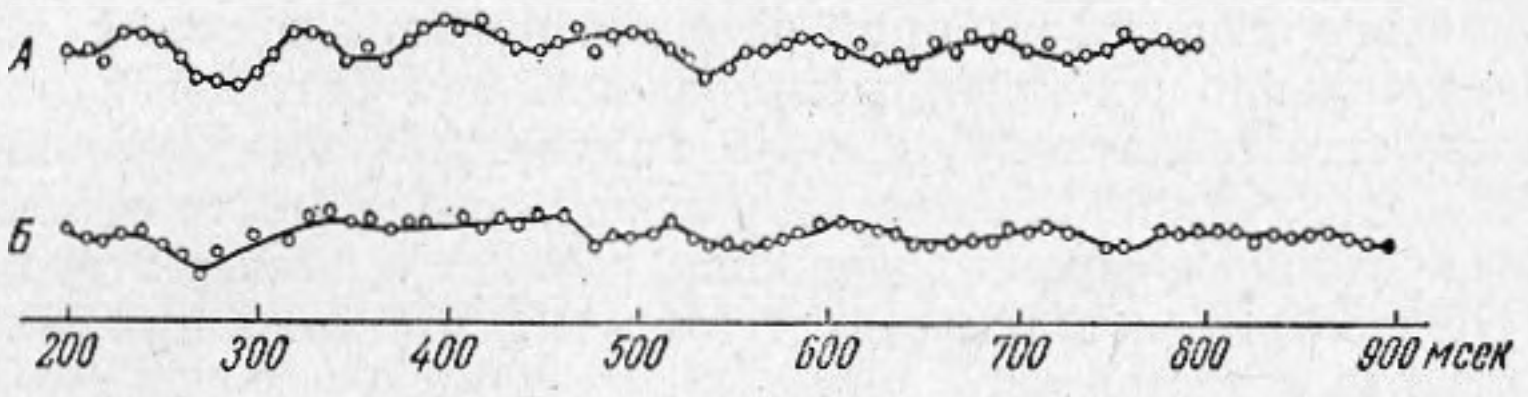


Фиг. 2

укорочение до 50 *мсек*, что согласуется с данными по влиянию укорочений на разборчивость английских гласных [8]. Особое положение звука *о* объясняется тем, что в русском языке он имеет дифтонгоидный характер [9]. Очевидно, что при записи естественного *о* испытуемые принимали решение о букве на основании накопления информации на отрезке звука, не меньшем 100 *мсек*, так как укорочение *о* до этой длительности уже вызывало удлинение времени задержки записи. Согласно данным по восприятию звуковых последовательностей [10], минимальный отрезок времени, с которым может быть связано одно решение, составляет примерно 100 *мсек*.

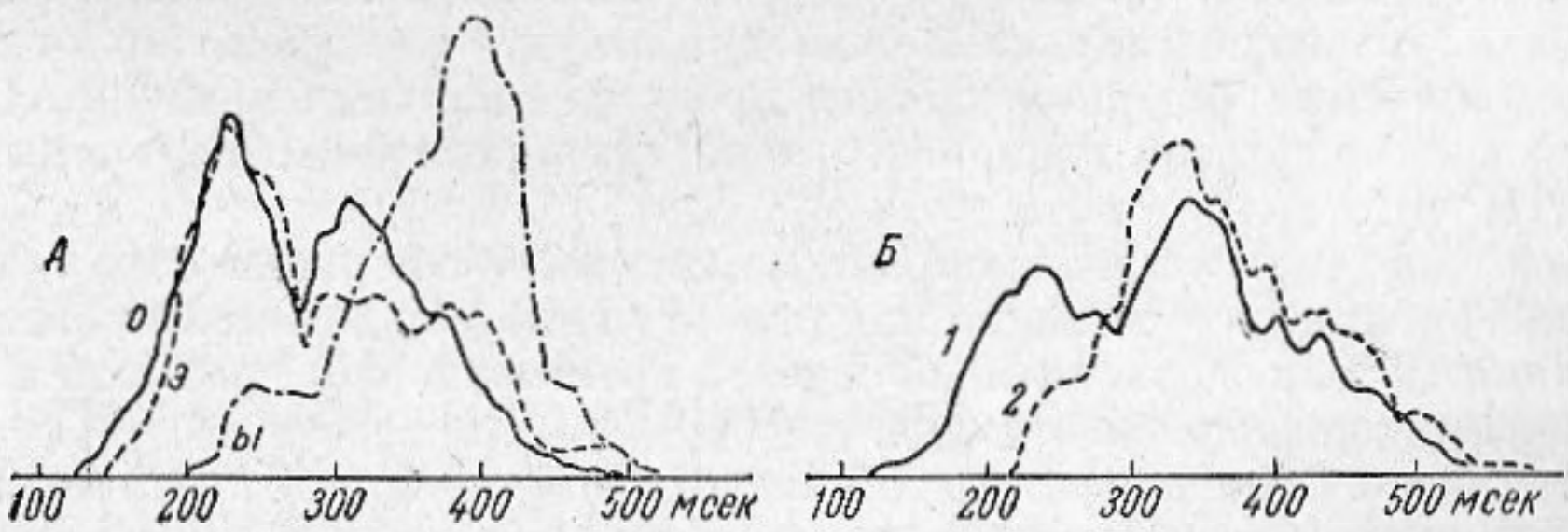
Итак, полученные данные заставляют прийти к выводу, что разница во временных задержках записи не может быть прямо связана с различным распределением информации на протяжении звуков — с их полезной длительностью. Это заставило искать каких-то иных механизмов связи

между свойствами сигнала и временем его распознавания. В поисках этих механизмов мы обратились к рассмотрению гистограмм распределения времен задержек для различных (естественных) гласных *. Гистограммы обнаружили несколько максимумов, причем создалось впечатление, что, независимо от характера гласных, максимумы частот реакций приходятся на одни и те же значения времени задержки с расстояниями между максимумами, соответствующими примерно 100 *мсек*. Разница между звуками



Фиг. 3

заключалась в разной относительной выраженности этих максимумов. Эти особенности распределений показались нам чрезвычайно интересными, так как, с одной стороны, периодичность в распределении времен задержек реакции, с периодом порядка 100 *мсек*, была уже описана в литературе [11, 12]; с другой стороны, такой периодичности можно было бы ожидать, исходя из гипотезы о периодическом, с частотой порядка 10 *гц*, сканировании (считывании) сигналов, приходящих в корковые проекционные области [13, 14].



Фиг. 4

Поэтому мы попытались прежде всего выделить предполагавшуюся периодическую составляющую в кривых распределения. Для этого сначала было произведено текущее усреднение кривой с периодом усреднения в 100 *мсек*. Для каждого значения времени задержки была вычислена средняя частота реакции на интервале $\pm$ 50 *мсек* от данного значения. Полученные таким образом значения ординат сглаженной кривой распределения были вычтены из соответствующих значений ординат кривой исходного распределения. Найденные величины разностей значений ординат приведены на фиг. 2, где на оси абсцисс отложено время задержки; нижние кривые — суммарные данные для всех гласных. Можно видеть, что кривые являются отчетливо периодическими и весьма близкими для разных звуков. У испытуемого *А* период несколько меньше (85 *мсек*), чем у *Б* (100 *мсек*). Результаты обработки данных, полученных в опытах с укороченными звуками (средняя длительность звука 15 *мсек*), показали, что при этом наблюдается та же периодичность, только менее четко выраженная (фиг. 3).

---

* Фактически для рассмотрения и дальнейшей обработки экспериментальные кривые были подвергнуты текущему усреднению по 50 *мсек*. Это было сделано для того, чтобы избавиться от пиков на 200, 250, 300, 350 и так далее *мсек*, введенных при обработке осциллограмм.

Очевидно, что периодичность в кривых распределения могла быть результатом того, что каждое распределение является сложным, образовано из нескольких простых распределений, средние которых отстоят друг от друга на величину периода. Воздействие характера звука может тогда рассматриваться как применение оператора, указывающего вероятности выборки из первого, второго и так далее простых распределений. Увеличение среднего времени задержки при переходе от одного гласного к другому, или при укорочении звука, должно быть в этом случае результатом повышения вероятности выборки из более поздних простых распределений. Сопоставление исходных распределений (см. фиг. 4, где даны распределения для естественного звука *а* — *1* и звука, укороченного до 15 *мсек*,—*2* для наблюдателя *Б*, и для естественных звуков *о*, *э*, *ы* для наблюдателя *А*, по абсциссе отложено время задержки), говорит в пользу этого предположения. Об этом же свидетельствует тот факт, что даже при сглаживании форма кривых остается сложной.

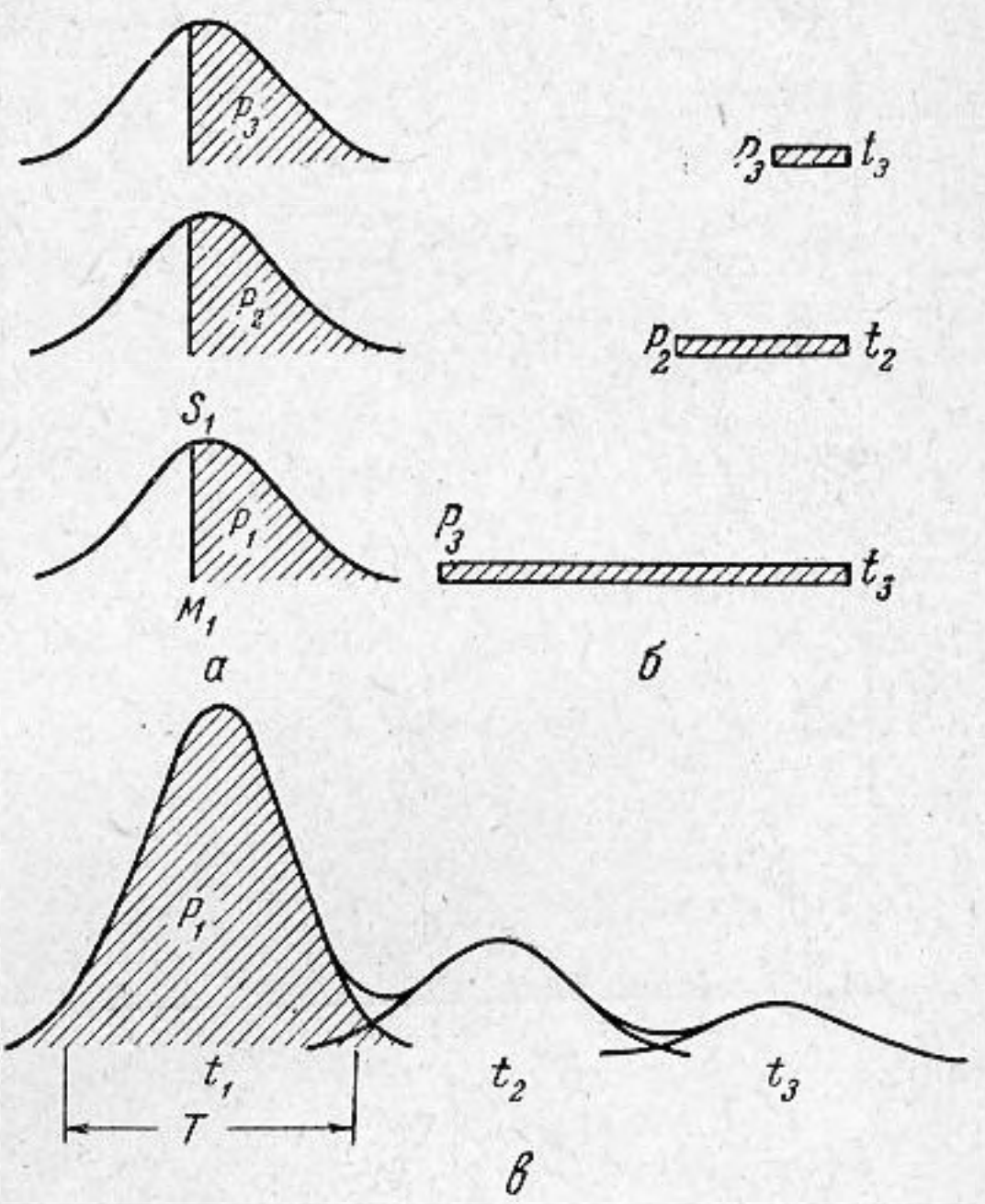
Периодичность в распределении времен задержек свидетельствует не только о том, что в нервной системе имеется какой-то периодический процесс, делающий наиболее вероятным возникновение реакции в определенные моменты времени. Она говорит также о том, что этот периодический процесс синхронизирован с сигналом. Если бы синхронизации не было, а сигнал попадал бы в случайную фазу цикла, то никакой периодичности в распределении времен задержек однократной реакции обнаружить не удалось.

Так как независимо от характера звука периодические составляющие кривых распределений совпадают друг с другом по фазе (фиг. 2), естественно считать, что таким неспецифическим синхронизирующим сигналом является включение звука. Можно думать, что электрофизиологическим проявлением этого периодического процесса являются медленные волны, следующие с частотой порядка 10 *гц* за так называемым первичным ответом [15, 16].

Получив данные, предсказываемые гипотезой о периодическом считывании информации, естественно принять эту гипотезу и использовать ее для создания модели, позволяющей связать время задержки реакции с информацией, содержащейся в звуке.

Принятая нами модель предполагает наличие блока кратковременной памяти, сохраняющей характеристики сигнала по полезным признакам, определяемым на интервале порядка 100 *мсек*, системы распознающих элементов и генератора считывания, подающего сигнал из памяти на распознающие элементы. Подача сигналов осуществляется в дискретные моменты времени $t_1, t_2, \ldots, t_n$, разделенные друг от друга на 80—100 *мсек*. Каждый распознающий элемент состоит из фильтра (набора характеристик), соответствующего эталону фонемы, и порогового индикатора. Сигнал на выходе фильтра пропорционален сходству между звуком и эталоном. Порог индикатора флюктуирует по нормальному закону со средним M_i. Ответ индикатора, назовем его решением, может возникать только в дискретные моменты времени, вероятность его на i-ом цикле определяется величиной сходства (S_k) относительно среднего значения порога. Если бы мы знали P_k (например, P_1 на фиг. 5, *а*), то тем самым определили бы в σ разность между величиной сходства и средним значением порога $S_k - M_k$ (например, $S_1 - M_1$ на фиг. 5, *а*), т. е. имели бы уже некоторую количественную меру сигнала. Однако величина P_k не может быть непосредственно определена экспериментально. Для нахождения ее из экспериментальных данных необходимо рассмотреть, какие отношения должны существовать между принятием внутреннего решения и внешней реакцией. Прежде всего очевидно, что внутренние решения могут быть повторными. Однако в наших и подобных нашим экспериментах человеку дается задание реагировать как можно быстрее, разрешается совершать на данный сигнал только одну реакцию и запрещается возможность исправлений.

Это заставляет считать, что, по условиям эксперимента, внешняя реакция соответствует только первому из решений, принятых в ответ на данный сигнал. Тогда распределение реакций должно соответствовать распределению вероятностей попадания первого принятого решения на первый, второй и так далее циклы считывания. Согласно модели, вероятность принятия решения на данном цикле определяется только величиной сходства звука с эталоном и параметрами распределения порогов на этом цикле. Она может рассматриваться независимо от вероятности принятия решения на других циклах. Иначе обстоит дело с вероятностью попадания первого решения на данный цикл (P_k); она естественно должна зависеть от вероятности принятия решения как на данном, так и на предыдущих циклах. Вероятность попадания первого решения на данный цикл равна вероятности принятия решения на данном цикле, помноженной на вероятность непринятия решения на всех предшествующих циклах (вероятность непопадания первого решения ни на один из предшествующих циклов).



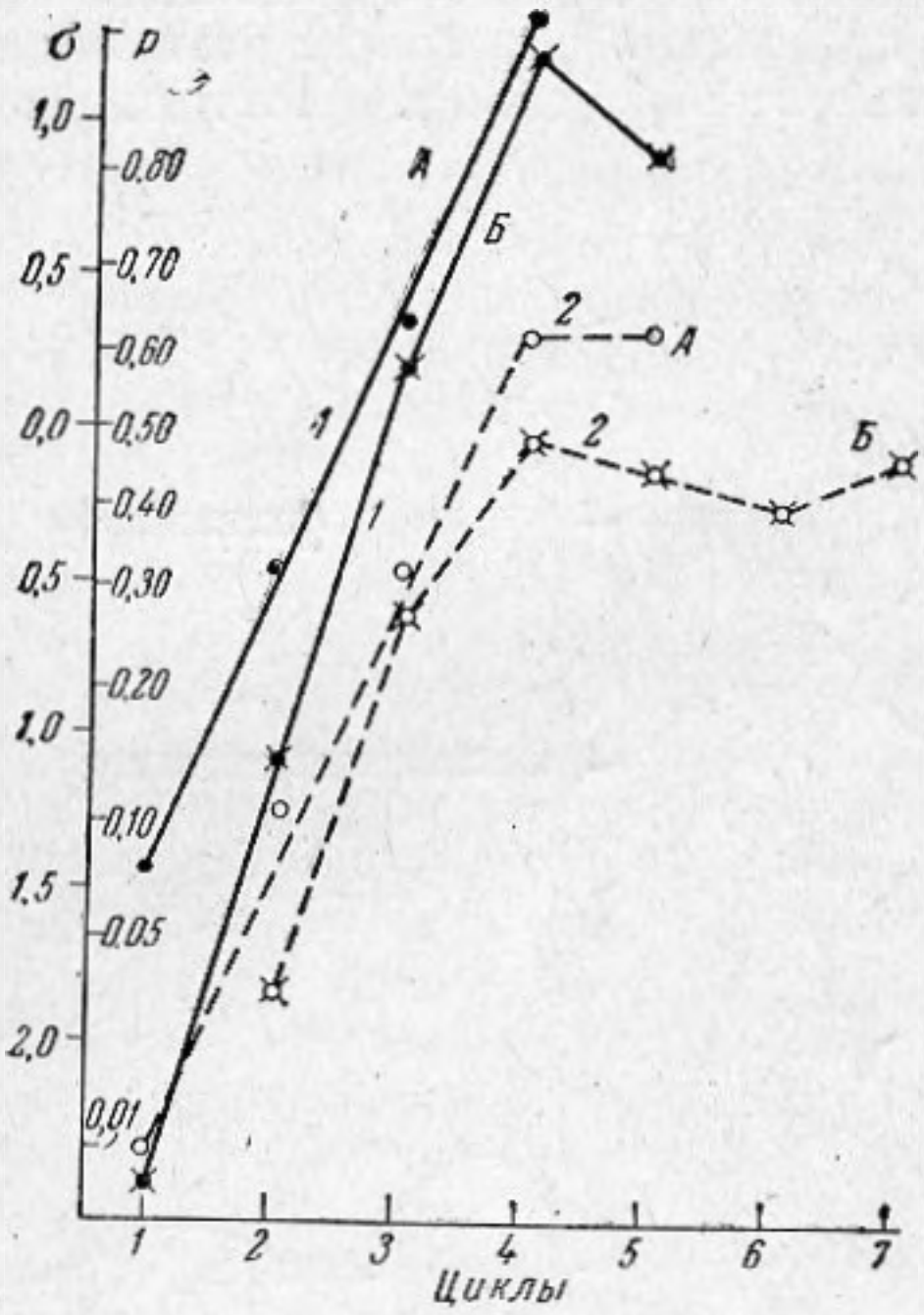
Фиг. 5

Так как $P_1 + P_2 + \ldots$ $\ldots P_n = 1{,}0$, то $P_k = p_k \left(1 - \sum_{i=k-1}^{i=1} P_i\right)$. Отсюда следует, что если бы мы знали распределение вероятностей попаданий первых решений на разные циклы, то могли бы вычислить и интересующие нас вероятности принятия решения, по крайней мере, на нескольких первых циклах.

Однако нужно сделать еще один шаг от распределения первых решений к распределению задержек реакции. Фиг. 5, *б* показывает, что распределение вероятностей попадания первого решения на различные моменты времени принимается дискретным, вероятность равна 0 всюду, кроме $t_1, t_2, \ldots, t_n$. Распределение реально измеренных времен задержки реакции должно приближаться к непрерывному из-за изменчивости поведения испытуемого (разное исходное состояние эффекторного органа, разное расстояние руки от бумаги и т. д.) и из-за случайных ошибок при измерении времени задержки экспериментатором. Естественно ожидать, что вместо дискретного распределения должна получиться сумма нормальных распределений со средними, отличающимися друг от друга на период считывания (фиг. 5, *в*). Для того чтобы восстановить из этого распределения или соответствующей ему экспериментально полученной гистограммы нужное нам дискретное распределение, можно принять за реакции, соответствующие решению в момент t_k, все реакции с задержками, находящимися в пределах данного цикла (интервал T на фиг. 5, *в*). Это и было нами сделано, однако полученные данные можно рассматривать только как ориентировочные, так как они могут быть сильно искаженными, если флюктуации реагирования и ошибки при обработке сравнимы с T.

Основываясь на рассмотренной модели, мы вычислили значения разности $S_k - M_k$ на различных циклах. Результаты показаны на фиг. 6, где приведены для наблюдателей *А* и *Б* средние данные по всем гласным. Кривые *1* относятся к естественной длительности гласных, *2* — к длитель-

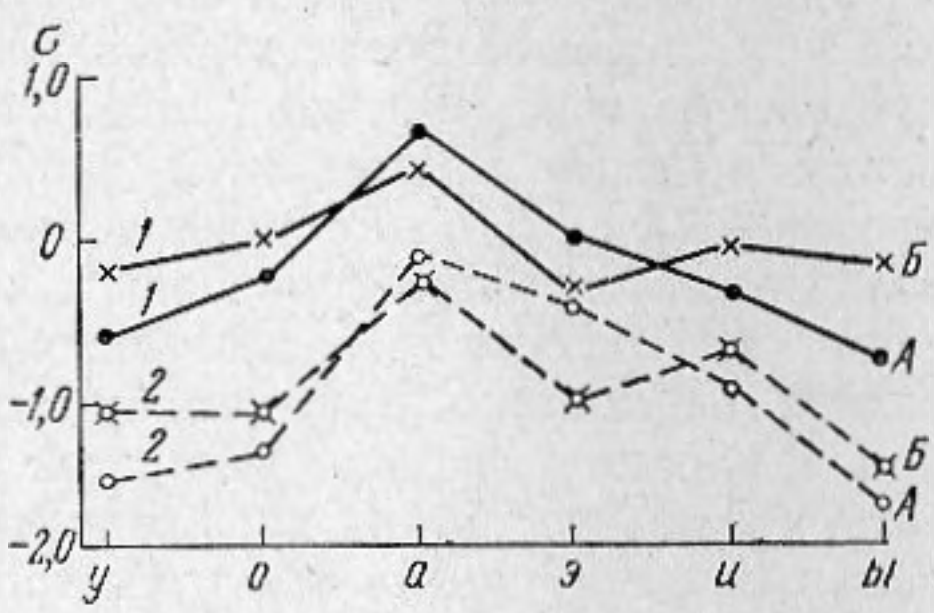
ности в 15 *мсек*. Можно видеть, что значение $S - M$ растет до 4-го цикла и затем остается примерно постоянным. Очевидно, что изменение $S - M$ может происходить как за счет увеличения сигнала (сходства между звуком и эталоном), так и за счет понижения порога. Так как у нас нет никаких оснований ожидать увеличения сходства от цикла к циклу и так как ход кривой, будучи аналогичным для естественных и укороченных звуков, различен у разных испытуемых, приходится думать, что мы имеем дело с изменением порога.



Фиг. 6

Понижение порога во времени может рассматриваться как весьма полезное приспособление, так как оно позволяет совместить большую надежность распознавания при отчетливых сигналах (большом сходстве с эталоном) со сравнительно малой задержкой распознавания неотчетливых сигналов. Кроме того, оно автоматически обеспечивает выбор, исключая необходимость сравнения друг с другом сходства предъявленного сигнала со всеми возможными фонемами.

Располагая кривыми *1* и *2* фиг. 6, мы можем использовать их в качестве уровня отсчета и вычислить относительную силу сходства для разных гласных (естественных и укороченных). Полученные данные приведены на фиг. 7 (обозначения те же, что и на фиг. 6). Можно видеть, что звуку *а* соответствует самое сильное сходство, звукам *у* и *ы* — самое слабое сходство. Для того чтобы понять эту разницу в силе сходства для разных гласных, необходимо допустить, что сила сходства зависит не только от степени близости между характеристиками сигнала и фонемы, по данному признаку, но и от числа признаков, позволяющих различить данную фонему от остальных возможных фонем. Это соответствует нашему интуитивному представлению о том, что узнать объект, отличающийся от других объектов по многим признакам, гораздо легче, чем узнать объект, отличающийся только по одному признаку. Если считать признаками фонем частоты первой и второй форманты, то фонема *а* оказывается в наиболее выгодном положении, так как она отличается от остальных фонем по обоим этим признакам [17, 18]. В наименее выгодном положении оказываются фонемы *у*, *ы*, так как они могут быть разделены только по частоте второй форманты, причем еще характеристики *и* , *ы* по этому признаку весьма близки друг к другу



Фиг. 7

Выводы

1. Исследовались временные задержки буквенной записи русских гласных. Обнаружено, что средние значения задержек записи составляют 300—400 *мсек*, т. е. примерно в два раза превышают задержки повторения. Делается вывод о невозможности текущего фонемного (графемного) распознавания каждого звука речи в речевом потоке.

2. Кривые распределения времен задержек записи обнаруживают отчетливую периодичность с максимумами, разделенными на 85—100 *мсек*. Положение максимумов по оси времен не зависит от характера звука; относительные амплитуды максимумов (и соответственно средние значения времен задержки) отчетливо зависят от характера гласного. Укорочение звуков приводит к уменьшению амплитуд первых максимумов и увеличению амплитуд (или появлению) более поздних максимумов.

3. Полученные данные подтверждают гипотезу о периодическом считывании информации в центральной нервной системе. Рассматривается модель, обеспечивающая зависимость задержки реакции от свойств сигнала.

ЛИТЕРАТУРА

1. E. C. Cherry. Some experiments on the recognition of speech with one and with two ears. J. Acoust. Soc. America, 1953, 25, 5, 975—979.
2. M. C. Saslov. Reaction time to consonant-vowel syllables in ensembles of various sizes. Quarterly Report, Laboratory of Electronics, M. I. T., July 1958, 143—144.
3. Л. А. Чистович, В. В. Алякринский, В. А. Абульян. Временные задержки при повторении слышимой речи. Вопр. психологии, 1960, 6, 1, 114—120.
4. Л. А. Чистович. Классификация звуков речи при их быстром повторении. Акуст. ж., 1960, 6, 3, 392—398.
5. Л. А. Чистович, Текущее распознавание речи человеком. Сб. «Машинный перевод и прикладная лингвистика» I МГПИИЯ, Москва, № 6, 7.
6. М. М. Бонгард. Моделирование процесса узнавания на цифровой счетной машине. Биофизика, 1961, 6, 2, 129—141.
7. D. E. Broadbent. Time to react. Advanc. Sci., 1960, 17, 65, 21—26.
8. R. Ahmend, R. Fatehchand. Effect of smple duration on the articulation of sounds in normal and clipped speech. J. Acoust. Soc. America, 1959, 31, 7, 1022—1029.
9. Л. Р. Зиндер. Общая фонетика, ЛГУ, 1960.
10. Л. А. Чистович. Восприятие звуковой последовательности. Биофизика, 1960, 5, 6, 671—676.
11. L. G. Augenstine. Evidences of periodicities in human task performance. Information Theory in Psychology, ed. H. Quastler, 1955, 208—226.
12. P. H. Venables. Periodicity in reaction time. Brit. J. Psychol., 1960, 51, 1, 37—43.
13. W. Pitts, W. S. Mc Culloch. How we know universals. The perception of auditory and visual forms. Bull. math. biophys., 1947, 9, 127.
14. W. G. Walter. The living brain. Gerald Duckworth et Co. L. T. D. L. 1953.
15. H. T. Chang. The repetitive discharges of cortico-thalamic reverberating circuit. J. Neurophysiol., 1950, 13, 3, 235—257.
16. M. R. Rosenzweig, W. A. Rosenblith. Responses to successive auditory stimuli at the cochlea and at auditory cortex. Psychol. Monogr., 1953, 67, 13, 1—26.
17. M. Halle. The sound pattern of Russian, Mouton et Co Gravenhage. 1959.
18. Л. А. Варшавский, Л. А. Чистович. Средние спектры русских гласных фонем. Пробл. физиол. акуст., 1959, 4, 181—186.

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

Поступила в редакцию
15 марта 1961 г.