АКУСТИЧЕСКИЙ ЖУРНАЛ

Том XX 1974 Вып. 3

УДК 534.75

НЕКОТОРЫЕ НОВЫЕ ПРЕДСТАВЛЕНИЯ ОТНОСИТЕЛЬНО СЛУХОВОГО АНАЛИЗА ЗВУКОВОГО ПОТОКА

С. Я. Жуков, М. Г. Жукова, Л. А. Чистович

Рассматриваются нейрофизиологические и психоакустические данные относительно обработки огибающей стимула и обнаружения моментов ее изменений. Описываются новые эксперименты, доказывающие, что слуховая система интерпретирует звуковой поток как последовательность временных отрезков между моментами изменений. Обнаружение момента изменения играет роль марки начала (конца) сегмента и используется для отбора информации о значении частот формант стимула в этот момент времени.

Классическая психоакустика рассматривает звуковой поток и слуховое ощущение как непрерывные функции времени. Это находит отражение не только в тех моделях, которые используются самой психоакустикой при интерпретации экспериментальных данных, но и в наиболее распространенных представлениях о процессе восприятия речи. Предполагается, что распознавание фонем основывается на классификации текущих значений спектра сигнала; классификатор вырабатывает новый символ (изменяет свое состояние), когда значения спектра перемещаются из одной фонемной области в другую.

Хотя принципиальные трудности, с которыми сталкивается применение такого подхода, широко известны (переслаивание фонемной информации), экспериментальные основания для пересмотра самих исходных представлений появились сравнительно недавно.

Было экспериментально показано, что для восприятия согласного в ряде случаев достаточно обнаружения определенного изменения в сигнале, причем это изменение может быть мгновенным (скачкообразное увеличение интенсивности стимула) или настолько быстрым (кратковременное прерывание звука, быстрое смещение частотного положения спектрального максимума), что его можно рассматривать как событие, точно локализованное во времени [1, 2].

Если слуховая система обнаруживает такие события, можно предположить, что она интерпретирует звуковой поток как последовательность временных отрезков между этими событиями и тем самым приобретает возможность использовать в качестве полезных признаков характеристики отрезков процесса (длительность отрезка, начальное, конечное и т. д. значения параметров процесса на отрезке).

Понятно, что принятие гипотезы слуховой сегментации потока существенно изменяет представления об операциях, из которых складывается обработка звуковой информации. На первый план выдвигаются такие задачи как изучение преобразований, обеспечивающих обнаружение моментов изменений стимула, и выяснение роли выходных сигналов этих обнаружителей в обработке данных и записи информации в память.

В настоящем обзоре будут рассмотрены вопросы обнаружения и интерпретации изменений в огибающей стимула.

Представление огибающей стимула на различных уровнях слуховой системы

При рассмотрении нейрофизиологических данных мы будем использовать два распространенных, но не общепринятых допущения. Первое из них состоит в том, что реакция всех нейронов по крайней мере трех первых уровней слуховой системы определяется характеристиками стимула в ограниченной области спектра — нейрон использует информацию, поступающую от одного или нескольких соседних фильтров улитки.

Второе весьма важное допущение заключается в том, что каналом, по которому передается информация, является не отдельный нейрон (нервное волокно), а группа нейронов, обладающих близкими характеристиками.

Понятие нейронного канала позволяет считать, что сигналом, передаваемым по каналу, является не дискретная последовательность стандартных по амплитуде нервных импульсов, а плотность импульсации, которую можно рассматривать как непрерывную (аналоговую) переменную. Если все нейроны, образующие данный канал, одинаковы, то плотность импульсации в момент t от начала стимула будет приближенно соответствовать вероятности возникновения импульса в этот момент времени, экспериментально определенной в результате многократных предъявлений стимула (постстимульные гистограммы).

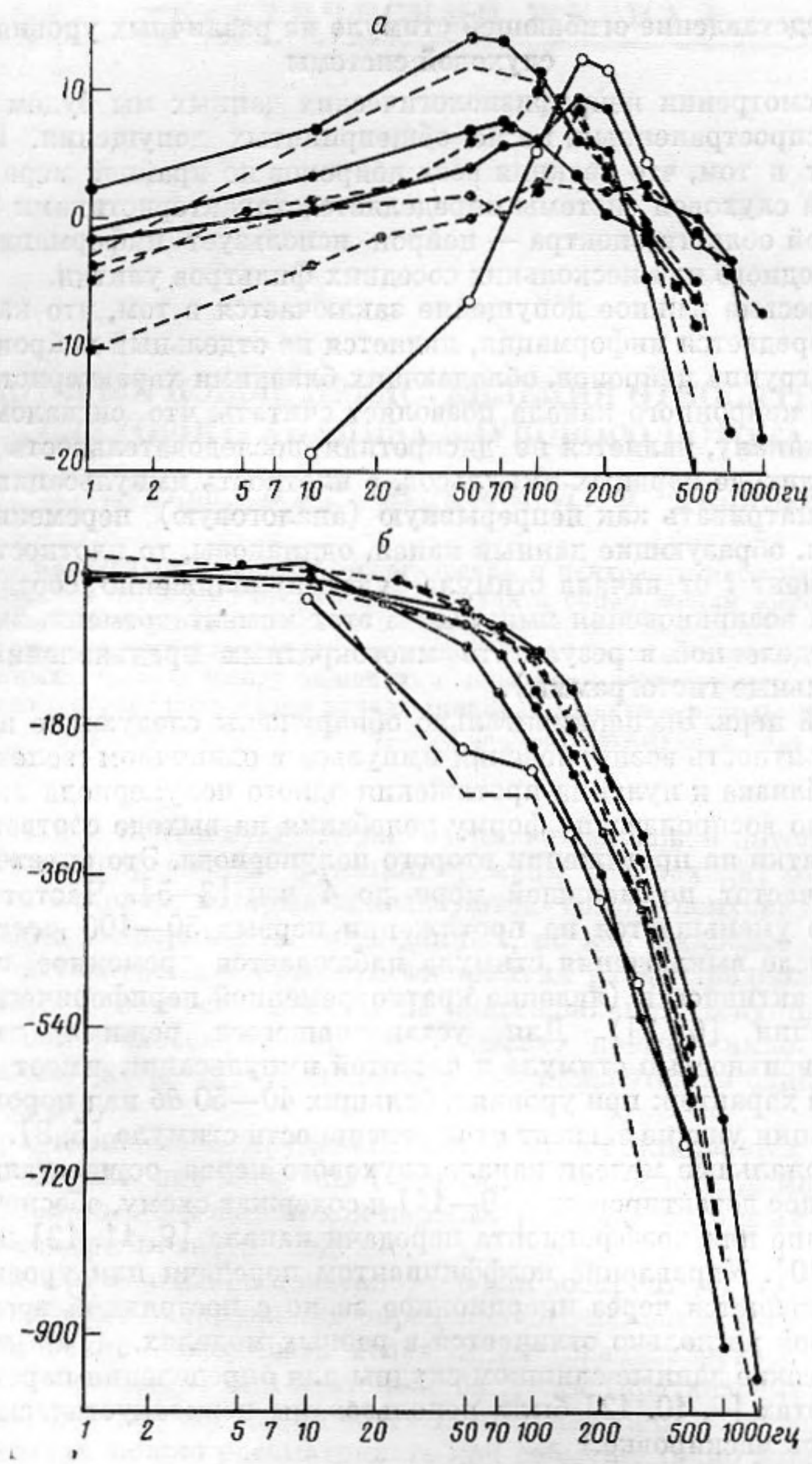
Слуховой нерв. Экспериментально обнаружены следующие важнейшие факты. Вероятность возникновения импульса в одиночном волокне слухового нерва близка к нулю на протяжении одного полупериода колебаний и приближенно воспроизводит форму колебания на выходе соответственного фильтра улитки на протяжении второго полупериода. Это справедливо для диапазона частот по меньшей мере до 4 *кгц* [3—5]. Частота импульсации резко уменьшается на протяжении первых 50—100 *мсек* действия стимула; после выключения стимула наблюдается временное подавление спонтанной активности (явление кратковременной периферической слуховой адаптации [6, 7]). Для установившегося режима зависимость между интенсивностью стимула и частотой импульсации имеет отчетливо нелинейный характер; при уровнях, больших 40—50 *дб* над порогом, частота импульсации уже не зависит от интенсивности стимула [5, 8].

Функциональные модели канала слухового нерва осуществляют однополупериодное детектирование [9—11] и содержат схему, обеспечивающую регулирование или коэффициента передачи канала [9, 11, 12] или уровня селекции [10]. Управление коэффициентом передачи или уровнем селекции осуществляется через инерционное звено с постоянной времени, значение которой несколько отличается в разных моделях. (Так как нейрофизиологические данные слишком скудны для определения параметров модели, в работах [9, 10, 12] были использованы психоакустические данные по остаточной маскировке.)

Действие этих моделей на огибающую стимула качественно напоминает пропускание огибающей стимула через дифференцирующую цепочку. При подаче на вход модели амплитудно-модулированного сигнала, при нижних модулирующих частотах глубина модуляции на выходе уменьшается с понижением модулирующей частоты. Верхний предел воспроизведения модуляции определяется шириной полосы соответственного фильтра улитки и должен быть разным в разных частотных каналах.

Второй слуховой нейрон (кохлеарные ядра). Постстимульные гистограммы значительной части нейронов этого уровня очень близки к таковым для волокон слухового нерва [13—15]. При действии амплитудно-модулированного сигнала огибающая плотности импульсации этих нейронов воспроизводит огибающую стимула вплоть до весьма высоких частот модуляции [16].

На постстимульных гистограммах второй группы нейронов наблюдается резко выраженный переходный процесс в момент включения (или также и выключения) стимула [13—15].



Фиг. 1

Функции передачи огибающей для ряда нейронов этой группы приведены на фиг. 1 [17]. Коэффициент передачи огибающей (по оси ординат на фиг. 1, *а*) определялся как отношение в децибелах глубины модуляции плотности импульсации на постстимульной гистограмме к глубине модуляции стимула. Измерения проводились при уровне интенсивности стимула, равной 20 *дб* над порогом нейрона.

Можно видеть, что для всех нейронов в некоторой области частот глубина модуляции плотности импульсации больше глубины модуляции стимула. Максимум находится (по всему материалу) в пределах от 50 до 200 *гц* [17]. В области более высоких частот коэффициент передачи резко уменьшается с частотой, фазовый сдвиг (фиг. 1, *б*) растет. Таким образом, канал действует как фильтр нижних частот, т. е. не передает колебаний плотности импульсации, соответствующих несущей частоте.

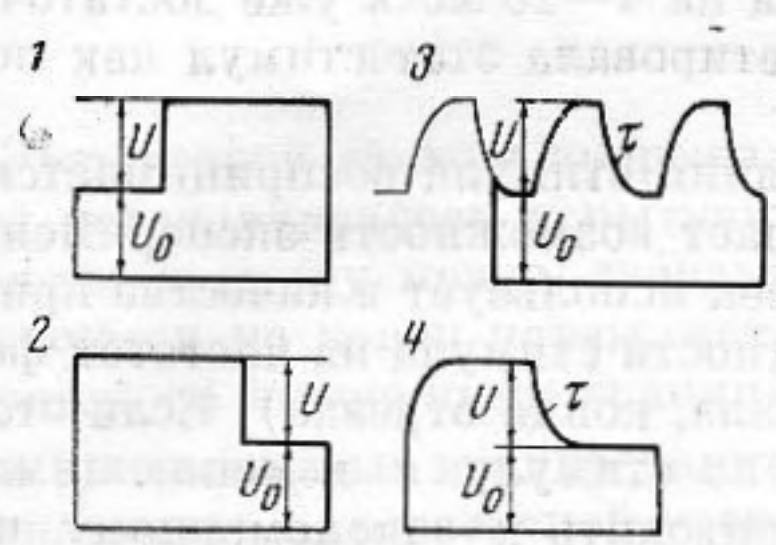
Характеристики, приведенные на фиг. 1, отражают, естественно, свойства не только нейронов кохлеарных ядер, но и всех периферических элементов канала (фильтр улитки, слуховой нерв). Несомненно, что подавление медленных изменений огибающей должно быть частично отнесено на счет явления периферической слуховой адаптации (см. выше). Существенно, что характеристики канала в целом таковы, что они обеспечивают выделение огибающей и подчеркивание ее быстрых изменений [17, 18].

Третий слуховой нейрон (заднее двухолмие). Примерно 30% нейронов этого уровня реагируют только в момент начала (или также окончания) звукового стимула; реакция состоит из одного—трех импульсов [19, 20]. Уже этот факт позволяет предполагать, что преобразование сигнала соответствует дифференцированию огибающей стимула. Имеющиеся данные косвенно подтверждают это предположение. Показано, что при увеличении постоянной времени включения стимула уровень интенсивности стимула, требующийся для вызова реакции, возрастает [20], нейрон не реагирует как при низких частотах амплитудной модуляции (до 20—30 *гц*), так и при частотах выше 100 *гц* [21]. К сожалению, прямых измерений функции передачи канала на уровне этих нейронов пока сделано не было.

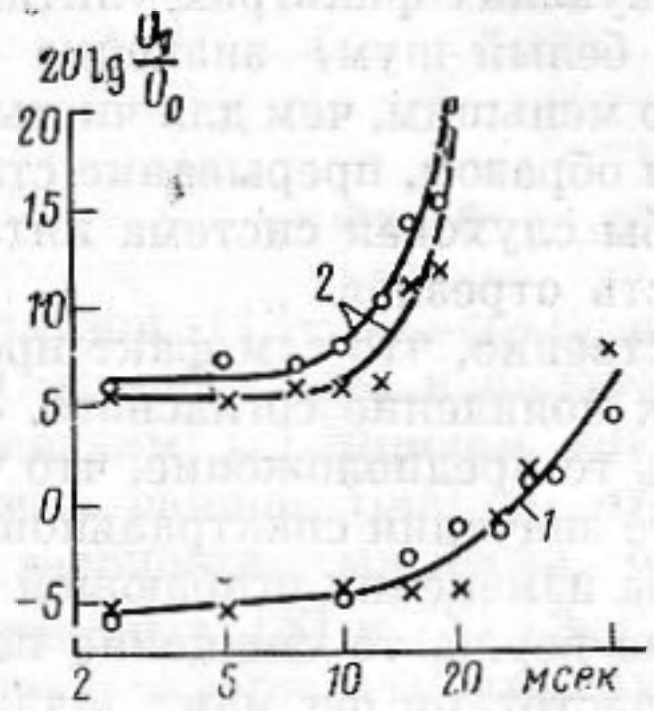
Обнаружение и фонетическая интерпретация изменений в огибающей стимула

На фиг. 2 приведены типы огибающих гласноподобного стимула, применявшихся при экспериментах.

Для стимулов типа 1 в работе [22] было показано, что если U больше некоторой пороговой величины, стимул восприимчивости как СГ (согласный — гласный) или ГСГ в зависимости от длительности первого сегмента [22]. В той же работе было установлено, что характер воспринимаемого согласного зависит от U ($|l|$ при малых приращениях, $|m|$ или $|n|$ при



Фиг. 2



Фиг. 3

больших). Далее, Жуковой было показано, что пороговые значения U/U_0, необходимые для обнаружения согласного, совпадают (в области средних уровней интенсивности) для стимулов типа 1 и типа 2.

В работе [23] применялись стимулы типа 3 с несколькими значениями U и τ; исследовалась фонемная идентификация стимулов. Полученные данные оказались совместимыми с предположением о том, что интерпретация стимула определяется скоростью изменения «мгновенной громкости», где «мгновенная громкость» пропорциональна квадратному корню из мгновенного значения огибающей звукового давления стимула.

В экспериментах Жуковой и Чистович применялись стимулы типа 4. При разных значениях τ испытуемый должен был установить значение U, соответствующее порогу появления фонетически неопределенного согласного (кривая *1* на фиг. 3) и порогу появления отчетливого $|m|$ (кривые *2*

на фиг. 3). Можно видеть, что при малых τ пороговое значение U не зависит от τ, в области больших τ увеличение τ приводит к увеличению U. Такого эффекта следовало бы ожидать, если бы для выделения изменений огибающей применялась обработка, соответствующая дифференцированию со сглаживанием. Так как наклон правой ветви кривых *2* фиг. 3 отчетливо больше 45°, следует заключить, что между мгновенными значениями огибающей стимула и мгновенными значениями «громкости» (плотности импульсации) существует нелинейная зависимость.

Полученные данные показывают, что человек обнаруживает как положительные, так и отрицательные приращения в огибающей стимула, причем переменная, на которую он реагирует, близка к скорости изменения огибающей.

Быстрые изменения огибающей как марки границ сегментов

Простейший способ создания быстрых изменений огибающей состоит в прерывании стимула.

В экспериментах Жукова и Чистович определялась минимальная величина паузы, при которой смещение паузы внутри стимула приводит к перемещению ударения с одного отрезка звука на другой. Так как положение ударения определяется отношением длин отрезков, перемещение ударения свидетельствует о том, что стимул воспринимается как сегментированный, у каждого отрезка обнаруживается начало и конец.

Зависимость минимальной величины паузы $t_п$ от частоты тона (общая длительность стимула 600 *мсек*, интенсивность — 40 *дб* над порогом ощущения) приведена на фиг. 4. Можно видеть, что значения пороговой паузы примерно одинаковы для частот выше 500 *гц*. Очевидно они определяются параметрами обработки огибающей в нервных отделах слуховой системы (см. выше). Рост пороговой паузы с понижением частоты в области низких частот должен быть отнесен на счет сглаживания огибающей стимула в соответствующих фильтрах улитки. Для широкополосных сигналов (гласный $|a|$, белый шум) значение пороговой паузы оказалось достоверно несколько меньшим, чем для чистых тонов (4—5 *мсек*).

Таким образом, прерывание стимула на 4—10 *мсек* уже достаточно для того, чтобы слуховая система интерпретировала этот стимул как последовательность отрезков.

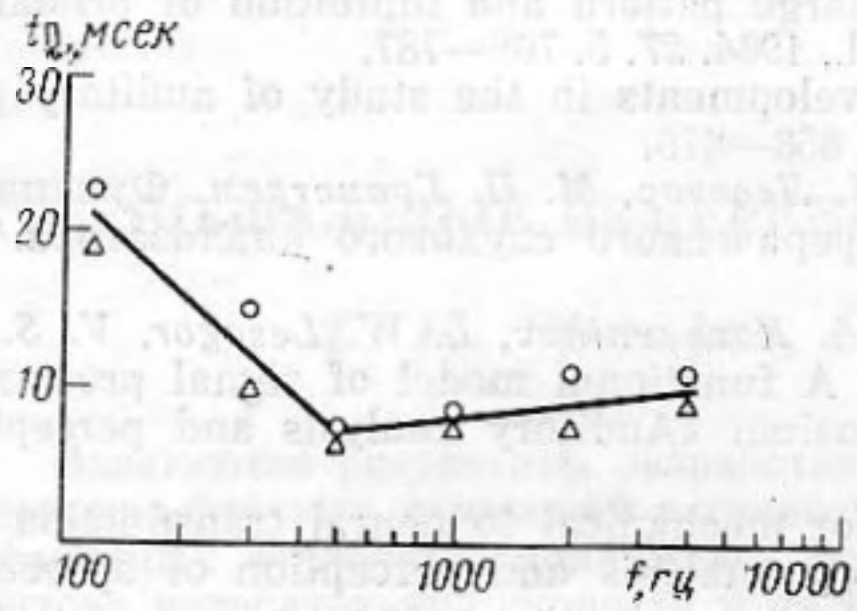
Существенно, что сам факт прерывания стимула воспринимается человеком как появление согласного. Это дает возможность экспериментально проверить то предположение, что человек использует в качестве признаков согласного значения спектральной плотности стимула на частотах формант в моменты изменения огибающей (начала, конца отрезка). Если это предположение верно, то смещение паузы по стимулу с изменяющимися во времени частотами формант должно приводить к закономерному изменению фонемного качества согласного.

В ряде работ было показано, что основным признаком, используемым для различия твердых и мягких согласных, является частота второй форманты [24—26]. Исходя из этих данных в экспериментах Жукова и Чистович был синтезирован стимул, формантный контур и огибающая которого схематически изображены на фиг. 5. Стимул прерывался на короткое время, что приводило к восприятию $|r|$. Задача испытуемого состояла в том, чтобы, смещая паузу вдоль стимула (изменяя t_1), найти положение паузы, соответствующее границе между твердым и мягким согласным (*iru* / *ir'u*).

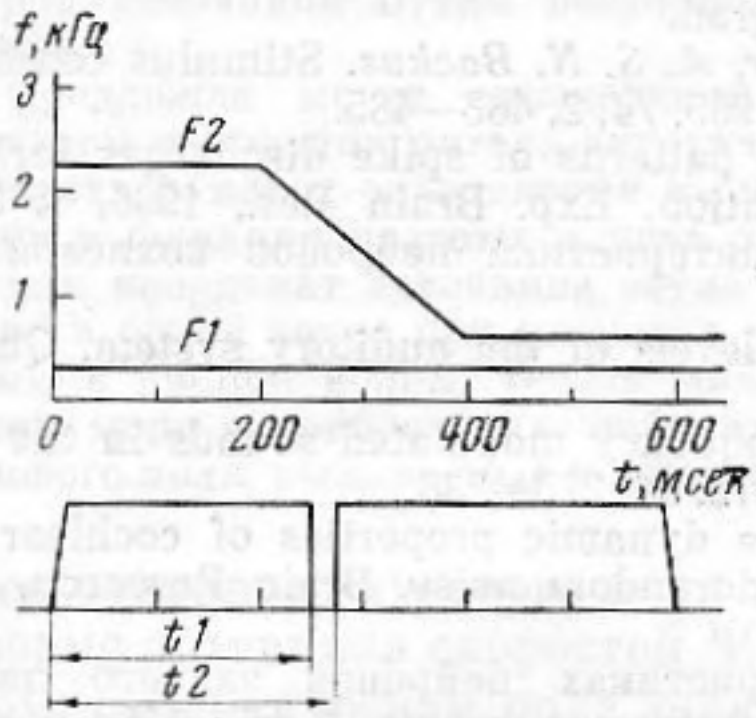
На фиг. 6, *а* показана зависимость стандартного отклонения установки t_1 (средние данные пяти испытуемых) от величины паузы; при паузах, больших 10 *мсек*, точность установки t_1 весьма велика.

На фиг. 6, *б* приведены средние значения устанавливаемых интервалов t_1 и t_2 в зависимости от величины паузы. Так как t_2 не зависит от величины паузы, можно заключить, что испытуемые ориентируются на значение F_2 в момент окончания паузы, т. е. в момент начала второго отрезка звука.

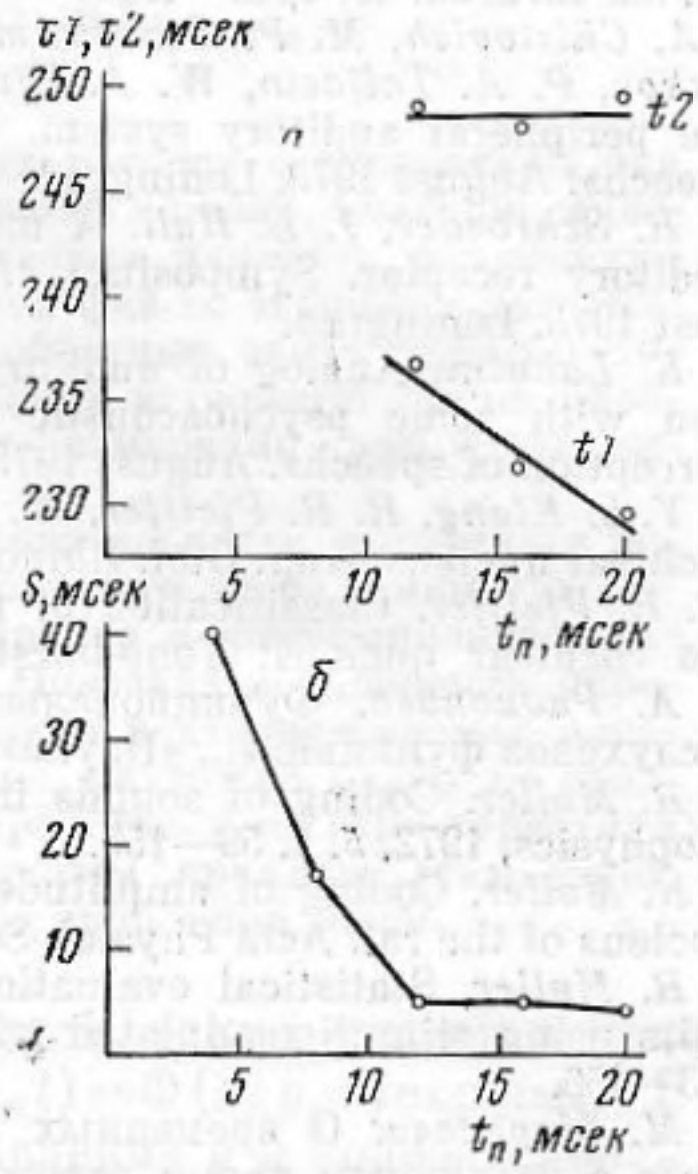
Окончательное доказательство предположения, что марка границы отрезка используется для отбора значений частот формант стимула в этот момент времени, было получено в экспериментах Жукова и Лисенко, использовавших дихотическую стимуляцию. На одно ухо (например, правое) подавался стимул с изменяющейся по времени F_2, который не прерывался. На левое ухо одновременно падавался второй стимул, в котором вторая форманта вообще отсутствовала, а частоты F_1 и F_3 не изменялись по вре-



Фиг. 4



Фиг. 5



Фиг. 6

мени. Этот второй стимул прерывался паузой (12—15 *мсек*), положение которой устанавливалось испытуемым. Оказалось, что испытуемые легко определяют границу между твердым и мягким $|r|$, причем значение F_2, приходящееся на конец паузы, находится в районе 1600 *гц*, что хорошо согласуется со значениями границы по твердости — мягкости, определенной для стационарных изолированных согласных $|S|$ и $|\check{S}|$ [24].

Итак, можно с достаточной уверенностью утверждать, что слуховая система интерпретирует звуковой поток как последовательность отрезков и использует в качестве полезной информации признаки этих отрезков. Так как признаки отрезков изменяющегося во времени процесса критично зависят от того, как определены границы отрезков, исследование слуховой сегментации и ее моделирование оказываются одной из важнейших промежуточных задач в изучении восприятия речевого потока.

ЛИТЕРАТУРА

1. *L. A. Chistovich, N. A. Fyodorova, D. M. Lissenko, M. G. Zhukova.* Auditory segmentation of acoustic flow and its possible role in speech processing. Symposium «Auditory analysis and perception of speech». August 1973. Leningrad.
2. *K. N. Stevens.* The potential role of property detectors in the perception of consonants. Symposium «Auditory analysis and perception of speech». August 1973, Leningrad.
3. *J. E. Rose, J. F. Brugge, D. J. Anderson, J. E. Hind.* Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol., 1967, *30*, 4, 769—793.

4. *J. F. Brugge, D. J. Anderson, J. E. Hind, J. E. Rose.* Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. J. Neurophysiol., 1969, *32*, 3, 386—401.
5. *J. E. Rose, J. E. Hind, D. J. Anderson, J. F. Brugge.* Some effects of stimilus intensity on response of auditory nerve fibers in the squirrel monkey. J. Neurophysiol., 1971, *34*, 4, 685—699.
6. *N. Y.-S. Kiang, T. Watanabe, E. C. Thomas, L. F. Clark.* Discharge pattern of single fibers in the cat's auditory nerve. Research Monograph № 35. Cambridge, Massachusetts The M. I. T. Press., 1965.
7. *M. Nomoto, N. Suga, Y. Katsuki.* Discharge pattern and inhibition of primary nerve fibers in the monkey. J. Neurophysiol., 1964, *27*, 5, 768—787.
8. *N. Y.-S. Kiang.* A survey of recent developments in the study of auditory physiology. Ann. Otol. Rhinol. Lar., 1968, *77*, 4, 656—676.
9. *Л. А. Чистович, И. А. Чистович, Л. В. Лесогор, М. П. Гранстрем.* Функциональная модель частотного канала периферического слухового анализатора. Физиол. ж., 1972, *58*, 10, 1548—1557.
10. *I. A. Chistovich, M. P. Granstrem, V. A. Kozhevnikov, L. W. Lesogor, V. S. Shupljakov, P. A. Taljasin, W. A. Tjulkov.* A functional model of signal processing in the peripheral auditory system. Symposium «Auditory analysis and perception of speech». August 1973, Leningrad.
11. *M. R. Schroeder, J. L. Hall.* A model for mechanical to neural transduction in the auditory receptor. Symposium «Auditory analysis and perception of speech». August 1973, Leningrad.
12. *V. K. Labutin.* Analog of auditory nerve response to acoustic signals and its relation with some psychoacoustic phenomena. Symposium «Auditory analysis and perception of speech». August 1973, Leningrad.
13. *N. Y.-S. Kiang, R. R. Pfeiffer, W. R. Warr, A. S. N. Backus.* Stimulus coding in the cochlear nucleus. Ann. Otol. Rhinol. Lar., 1965, *74*, 2, 463—485.
14. *R. R. Pfeiffer.* Classification of response patterns of spike discharges for units in the cochlear nucleus: Tone-burst stimulation. Exp. Brain Res., 1966, 1, 220—235.
15. *Е. А. Радионова.* Функциональная характеристика нейронов кохлеарных ядер и слуховая функция. Л., «Наука», 1971.
16. *A. R. Møller.* Coding of sounds in lower levels of the auditory system. Quart. Rev. Biophysics, 1972, *5*, 1, 59—155.
17. *A. R. Møller.* Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat. Acta Physiol. Scand., 1972, *86*, 223—238.
18. *A. R. Møller*, Statistical evaluation of the dynamic properties of cochlear nucleus units using stimuli modulated with pseudorandom noise. Brain Research, 1973, *57*, 443—456.
19. *А. М. Марусева.* О временных характеристиках нейронов заднего двухолмия крыс с различным типом ответов на звуковые сигналы. В сб.; Механизмы слуха. Л., «Наука», 1967, 50—62.
20. *И. А. Вартанян, В. Н. Снетков.* Зависимость импульсных реакций нейронов задних холмов от постоянной времени звукового сигнала. Физиол. ж. 1970, *56*, 5, 696—706.
21. *И. А. Вартанян.* Импульсная активность нейронов заднего двухолмия на звуковые сигналы, модулированные по амплитуде. Физиол., ж., 1969, *55*, 3, 273—282.
22. *Ю. И. Кузьмин, Д. М. Лисенко.* Фонетическая интерпретация стимулов с резкими изменениями интенсивности. В сб.: Анализ речевых сигналов человеком. Л., «Наука», 1971, 83—91.
23. *С. Я. Жуков, В. А. Кожевников, Ю. И. Кузьмин, Л. А. Чистович.* Образование слога при амплитудной модуляции гласноподобного звука. В сб.: Вопросы теории и методов исследования восприятия речевых сигналов, № 2, Информационные материалы Объединенного совета «Физиология человека и животных», вып. 32, Л., 1971, 36—44.
24. *В. С. Щупляков.* Акустический признак восприятия мягкости стационарных шумных согласных |s| и |š|. Тезисы докладов VI Всес. акуст. конф. М., 1968, 3Ш10
25. *В. С. Щупляков, Г. Фант, А. Серпа-Лейтао.* Акустический признак твердых и мягких согласных русского языка в связной речи (спектрографическое исследование). В сб.: Анализ речевых сигналов человеком. Л., «Наука», 1971, 57–65.
26. *M. Derkach.* Acoustical cues of softhess in Russian syllables and their application in the automatical speech recognition. Symposium «Auditory analysis and perception of speech». August 1973, Leningrad.

Институт физиологии им. И. П. Павлова
Академии наук СССР

Поступила
21 января 1974 г.