

КОДИРОВАНИЕ РЕЧЕВЫХ СИГНАЛОВ ДЛЯ ЦЕЛЕЙ ЭЛЕКТРОДНОГО ПРОТЕЗИРОВАНИЯ СЛУХА

Л.В. ЛЕСОГОР

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

РЕЗЮМЕ

Описаны психоакустические эксперименты с восприятием кодированных речевых сигналов нормально слышащими испытуемыми. Кодирование осуществлялось таким образом, что исходный сигнал преобразовывался в последовательность коротких биполярных импульсов, модулированных по амплитуде огибающей речевого сигнала.

Исследовалось влияние различных параметров кодирования на словесную разборчивость.

ВВЕДЕНИЕ

Электродное протезирование слуха или кохлеарная имплантация является одним из новейших направлений реабилитации слуха при полной глухоте. Основная задача, которая стоит перед исследователями, занимающимися электродным протезированием состоит в том, чтобы осуществить такое преобразование речевого сигнала в электрические стимулы, которое с одной стороны сохраняло бы как можно больше информации о речевом сигнале, с другой — удовлетворяло тем условиям, которые возникают в связи со спецификой электрического возбуждения слухового нерва.

Все имеющиеся в настоящее время системы электродного протезирования разделяют на одноканальные и многоканальные. Принципиальное отличие многоканальной системы от одноканальной состоит в том, что многоканальные системы протезирования, имеют

несколько параллельных каналов, разделяющих акустический сигнал на спектральные полосы посредством фильтров, и таким образом могут обеспечить изображение речевых сигналов, более близкое к естественному. При одноканальном же стимулировании необходимо произвести такое кодирование акустического сигнала, которое позволит выделить из него признаки, содержащиеся во временной картине [1]. Одноканальные системы обладают рядом существенных достоинств, и именно поэтому, в настоящее время они получили наибольшее распространение. Основные преимущества заключаются в следующем: 1. Наличие одного канала и, следовательно, одного электрода, значительно упрощает конструкцию и, что особенно важно, имплантацию кохлеарных протезов. 2. Только одноканальная стимуляция имеет смысл в тех случаях, когда у больного остается небольшое количество сохранившихся (непораженных) волокон слухового нерва. 3. Временная структура стимулирующего сигнала является гораздо информативнее, чем это предполагалось до последнего времени; с её помощью может передаваться не только просодическая информация, но и информация о фонемах, а также можно достичь удовлетворительного распознавания как отдельных согласных и односложных слов, так и слитной речи [4]. Более эффективными являются такие одноканальные системы электродного протезирования, в которых преобразование осуществляется таким образом, что информация о

частотном составе речевого сигнала, кодируется величиной межимпульсных интервалов, а информация об интенсивности — либо длительностью, либо амплитудой импульсов [5].

В настоящее время разработано и испытывается много типов речевых преобразователей, проводятся их сравнительные исследования, однако до сих пор еще нет четких данных, касающихся выбора наиболее оптимальных параметров кодирования: частоты среза фильтра, длительности биполярных импульсов и т.д. Это можно объяснить тем, что исследования, проводимые на пациентах с имплантируемыми электродами усложняются целым рядом практических трудностей: а) число возможных пациентов мало (1 или 2); б) имплантируемое устройство налагает ограничения на сигнал стимуляции; в) результаты зависят от таких неконтролируемых факторов, связанных с патологией улитки, как число и характеристическая частота уцелевших слуховых нейронов. Поэтому нами была предпринята попытка решить эту задачу с помощью испытуемых, обладающих нормальным слухом, исходя из того, что оптимальность параметров выбранная испытуемыми с нормальным слухом будет также действительна и для глухих пациентов. Известно, что использование нормальнослышащих испытуемых для сравнительного анализа схем речевого кодирования в работе [2] дало хорошие результаты.

ОПИСАНИЕ КОДИРУЮЩЕГО УСТРОЙСТВА И ПРОЦЕДУРА ИЗМЕРЕНИЙ

На основании вышеизложенного нами было разработано устройство, позволяющее кодировать как речевой, так и любой акустический или электрический сигнал одним, как нам кажется, из наиболее оптимальных способов, при котором исходный сигнал преобразуется в последовательность коротких биполярных импульсов, модулированных по амплитуде огибающей речевого сигнала.

Преобразователь выполнен в виде отдельного устройства, на вход которого сигналы могут подаваться либо с микрофона, либо от любого источника электрических сигналов. В устройстве имеется фильтр пропускания низких частот (ФНЧ), частоты среза которого устанавливаются переключателем и принимают значения от 0.5 кГц до 3.0 кГц ступенями через 0.5 кГц. Динамический диапазон амплитуды выходных импульсов определяется амплитудной характеристикой входящего в него логарифмического усилителя, который обладает компрессией 1:10. Амплитудная огибающая переменного сигнала выделяется детектором и затем сглаживается с помощью интегратора. Формирование импульсов осуществляется посредством триггера Шмитта и одностабильных мультивибраторов. На выходе, таким образом, получают амплитудно-модулированные биполярные импульсы заданной длительности, моменты появления которых соответствуют моментам пересечения нуля (в положительной фазе) сигнала с выхода ФНЧ.

Процедура измерения состояла в следующем: записанные на грампластинку слова с речевого аудиометра АР-03 поступали на кодирующее устройство, телефоны. Прослушивание производилось в звукоизолированной камере, моноурально, испытуемыми с нормальным слухом. Использовались специально составленные таблицы слов Г.И.Гринберга, Л.Р.Зиндера и Л.В.Нейман. Были проведены 4 серии экспериментов по определению наиболее оптимальных значений следующих параметров электрических стимулов: частоты среза ФНЧ, f_c ; длительности биполярных импульсов, $t_{и}$; постоянной времени интегрирования, τ ; интенсивности подаваемых стимулов, L : С этой целью были сняты зависимости разборчивости речевых тест-таблиц от значения этих параметров. Разборчивость определялась по количеству правильно воспринятых и записанных испытуемыми слов (из 60 предъявленных) и выражалась в процентах.

$$p = \frac{1}{m \cdot n} \sum_{k=1}^{k=n} \sum_{i=1}^{i=m} p_{ik} [\%]$$

где p_{ik} - процент правильно принятых слов i -ым аудитором в k -ой таблице, m - число испытуемых, n - число прослушанных таблиц.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Данные, полученные в экспериментах по определению разборчивости речевого теста в зависимости от частоты среза фильтра нижних частот, (при постоянной времени сглаживания 5 мсек и длительности биполярных импульсов 100 и 500 мсек), приведены на рис. 1.

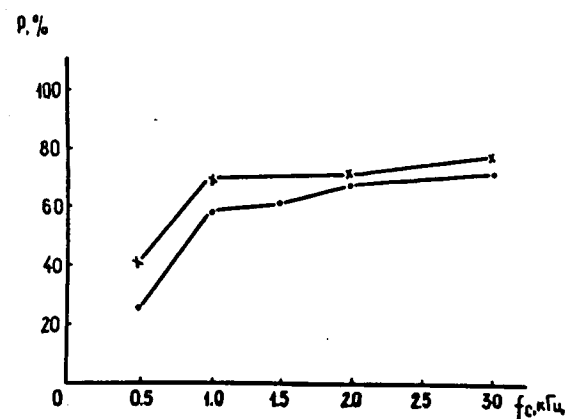


Рис. 1. Зависимость словесной разборчивости от частоты среза ФНЧ.

По оси ординат - разборчивость в процентах; по оси абсцисс - частота в кГц (усредненные данные для 5 испытуемых). X - $t_n = 100$ мсек; o - $t_n = 500$ мсек $\tau = 5$ мсек.

Можно видеть, что разборчивость резко возрастает с расширением полосы пропускания сигнала до 1,0 кГц; при дальнейшем увеличении частоты среза разборчивость меняется незначительно. Исследования нейронных

ответов при электрической стимуляции также указывают на ограниченный частотный диапазон, не превышающий 1 кГц, при котором сохраняется информация о периоде колебаний стимулов. Это позволяет предположить, что если нормальнослышающий испытуемый способен различать речевые стимулы, прошедшие НЧ фильтрацию с частотой среза 1,0 - 1,5 кГц, то и пациент с одноканальной стимуляцией внутреннего уха, диапазон воспринимаемых частот которого ограничен такой же частотной областью сможет распознавать речевые стимулы, используя такие параметры, как частоту основного тона и частоту 1-й и 2-й формант, лежащих в пределах этой области.

Во второй серии экспериментов исследовалось влияние постоянной времени сглаживания амплитудной огибающей, которая принимала значения, равные 5, 20 и 50 мсек, на разборчивость речевого теста при постоянных значениях частоты среза (1,5 кГц) и длительности импульсов (500 мсек). Результаты эксперимента показали, что постоянная времени сглаживания очень незначительно влияет на разборчивость. Анализ ошибок ответов указывает на то, что увеличение постоянной времени сглаживания влияет на восприятие лишь начальной фонемы в слове. Поэтому при $\tau = 50$ мсек увеличивается число ошибок в распознавании начальной фонемы. При значениях $\tau = 5$ мсек качество звучания ухудшается за счет восприятия щелчка.

На основании вышесказанного можно сделать вывод о том, что наиболее оптимальными с точки зрения правильного распознавания, являются значения постоянной времени сглаживания 10-20 мсек.

В третьей серии экспериментов определялась зависимость словесной разборчивости от длительности биполярных импульсов. Длительность импульсов менялась в пределах от 100 мсек до 1000 мсек. Данные были получены для трех испытуемых. Исходя из результатов предыдущих серий экспериментов,

частота среза ФНЧ была выбрана равной 1,5 кГц, постоянная времени сглаживания амплитудной огибающей исходного сигнала 10 мс. Из эксперимента следует, что с увеличением длительности биполярных импульсов до 500 мсек, процент правильного восприятия монотонно возрастает, а затем остается постоянным, равным $\approx 70\%$. Наши результаты хорошо согласуются с данными, полученными у пациентов с имплантированными электродами [3].

Из вышесказанного следует, что оптимальная длительность биполярных импульсов составляет около 500 мсек.

В четвертой серии экспериментов измерялась зависимость словесной разборчивости от уровня подаваемых речевых стимулов. Т.е. определялось, существует ли какой-то наиболее оптимальный уровень интенсивности речевых стимулов, при котором разборчивость получается наибольшей. Уровень звукового давления измерялся в дБ над порогом слышимости отдельно для каждого испытуемого. Как можно видеть из рисунка 2 даже при низком уровне интенсивности стимулов (10 дБ) разборчивость достаточно высока; для 2-х испытуемых она превышает 50% и намного выше начального уровня разборчивости (20%) для третьего испытуемого.

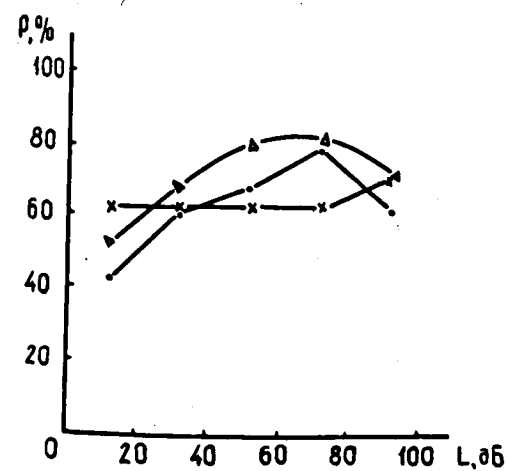


Рис. 2. Зависимость словесной разборчивости от уровня подаваемых речевых стиму-

лов для трех испытуемых.

Δ - исп. А; x - исп. Б; o - исп. В при $f_c = 1,5$ кГц и $\tau = 10$ мсек. Наиболее оптимальным уровнем, при котором разборчивость достигает 82% (исп. А и Б), оказался уровень 70 дБ. При дальнейшем увеличении интенсивности сигнала происходит некоторое ухудшение разборчивости, вызванное слишком высоким уровнем интенсивности стимулов, близким к болевому порогу. Таким образом, результаты данной работы позволили посредством психоакустических экспериментов произвести предварительный выбор оптимальных параметров электрических сигналов, что является необходимым шагом на данном этапе исследований в области электродного протезирования, т.к. позволит более обоснованно подходить к созданию аппаратуры для кодирования речи и сократить время постоперационного тестирования больных с имплантируемыми электродами.

ЛИТЕРАТУРА

1. Люблинская В.В. Слуховое восприятие у человека при электрическом раздражении слуховой системы. Раздражение периферических отделов слуховой системы. В кн.: Электродное протезирование слуха. Л.: "Наука", 1984, стр. 147-171.
2. Blamey P.Y., Martin L.F.A., Clark G.M. A comparison of three speech coding strategies using an acoustic model of a cochlear implant. - J. Acoust. Soc. Amer., 1985, v.77 (1), p. 209-217.
3. Burian K., Hochmair E., Hochmair-Desoyer I.J., Lessel M.B. Designing of the experience with multichannel cochlear implants. - Acta otolaryngol., 1979, v.86, p. 190-195.
4. Hochmair-Desoyer I.J., Hochmair E.S., Fischer R.E., Burian K. Cochlear prosthesis in use: Recent speech comprehension results. - Arch. Oto-Rhino-Laryngol., 1980, v.229, p. 81-98.