

УДК 62.506.2

С. А. УСЕНКО

К ВОПРОСУ О ДИСКРЕТНОСТИ СЛУХОВОЙ ИНФОРМАЦИИ

В настоящее время в области распознавания, синтеза и передачи звукового сигнала на расстояние возник ряд технических задач, связанных с созданием эффективных устройств кодирования речевого сигнала. Одним из важнейших источников совершенствования этих устройств является изучение органа слуха человека, представляющего собой естественный образец ввода информации и существенно превосходящий имеющиеся на сегодняшний день технического устройства, связанные с переработкой акустической информации.

Задачей данной работы является попытка показать на основе психоакустических экспериментов, что количество информации, извлекаемое слуховым аппаратом человека из звуковой осциллограммы, является конечным. Эта конечная информация, представленная в дискретном виде, может быть введена в цифровые устройства вычислительной машины и обработана на дискретном уровне.

Обращаясь к исследованиям в области психофизики, мы находим подтверждение того положения, что информация, воспринимаемая органами чувств, может быть задана в дискретном виде. Это положение вытекает прежде всего из того факта, что любой сенсорный анализатор можно рассматривать как неидеальный прибор, воспринимающий и преобразующий непрерывную информацию и не реагирующий на слишком малые изменения характеристик преобразуемой информации.

В работе [1] на примере зрительного анализатора приводятся виды ограничений, присущие любому приемнику информации и позволяющие рассматривать эту информацию как конечную. К таким ограничениям автор относит разрешающую способность прибора, его ограниченную чувствительность, а также ограничение полосы пропускания. Рассмотрим последний вид ограничений подробнее, поскольку он имеет прямое отношение к вопросам, поставленным в настоящей статье.

В силу известной теоремы Котельникова [2] ограничение полосы пропускания эквивалентно тому, что при передаче информации вместо обычного непрерывного времени вводится условное дискретное время, соседние моменты которого отличаются друг от друга на весьма малый отрезок времени. В качестве такого отрезка времени выбирается максимальный отрезок, в течение которого рассматриваемый прибор оказывается неспособным различить изменения величины несущей информации. Поскольку теорема Котельникова применима исключительно к линейным системам, а процесс преобразования информации слуховым аппаратом человека носит нелинейный характер, мы вынуждены искать новые доказательства факта конечности акустической информации. Таким доказательством может служить обобщенный закон Тальбота для слуховых ощущений, описанный в работе [3], который позволяет рассматривать орган слуха как своеобразный фильтр, ограничивающий объем информации, которая поступает из внешнего мира в мозг человека. Заметим, что закон Тальбота свидетельствует о том, что некоторые существенно различные входные сигналы преобразуются органом слуха в одинаковые выходные сигналы. Указанные опыты [см. 3], очевидно, не исчерпывают всего класса сигналов, удовлетворяющих этому закону.

Для экспериментального доказательства дискретности слуховой информации необходимо предъявить испытуемому произвольную осциллограмму звукового сигнала $x(t)$ и поставить ей в соответствие дискретный сигнал $x_n(t)$, выбранный таким образом, чтобы

$$\int_{t_1}^{t_2} x(t) dt = \int_{t_1}^{t_2} x_n(t) dt. \quad (1)$$

Согласно закону Тальбота ощущения, формируемые этими сигналами, должны совпадать. Поскольку $x_n(t)$ в равенстве (1) может быть задан неоднозначно, возникает вопрос, какой вид дискретного сигнала необходимо использовать для достижения целей эксперимента. С одной стороны, этот сигнал должен удовлетворять равенству (1), с другой — целесообразно представить его серией стандартных импульсов, следующих с частотой, превышающей критическую частоту слития прерывистого звука в непрерывный. Как показывают эксперименты [3], $f_{кр} \geq 30$ кГц. Информация, содержащаяся в сигнале и сформированная серией стандартных импульсов, может быть легко представлена двоичным кодом. Очевидно, что такой дискретный сигнал $x_n(t)$ можно получить путем частотно-импульсной модуляции сигнала $x(t)$.

Обращаясь к вопросам генерации импульсов в нейронных сетях, уместно заметить, что именно такой способ кодирования имеет место в слуховом анализаторе человека [4]. Возбуждение нейрона происходит при условии, что интенсивность воздействующего на нейрон раздражителя превышает некоторую величину — нижний порог чувствительности, или абсолютный порог. По мере увеличения интенсивности раздражителя возрастает количество

или частота импульсов, генерацией которых нейрон откликается на данное раздражение.

Опишем аналитически такой вид преобразования. Допустим, на вход слухового аппарата поступает какой-то акустический стимул. В предположении дискретности выходного сигнала слухового аппарата это преобразование будет иметь вид, показанный на временной диаграмме (рис. 1), т. е. на выходе слухового аппарата появляется стандартный сигнал, например какой-либо импульс, когда

$$\int_0^t f(\tau) d\tau = E, \quad (2)$$

где E — энергия порога срабатывания.

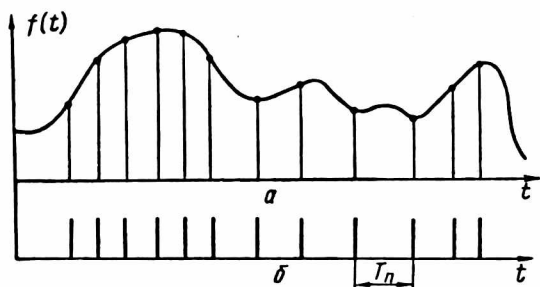


Рис. 1. Временная диаграмма работы преобразователя: a — непрерывный входной сигнал; b — преобразованный дискретный сигнал.

Таким образом, информация в выходном сигнале будет переноситься, в основном, интервалом времени между импульсами T_n .

Найдем зависимость $T_n(t)$ от $f(t)$, т. е. функцию преобразования

$$\int_t^{t+T_n} f(\tau) d\tau = \Phi[t + T_n(t)] - \Phi(t) = E, \quad \text{где } \Phi[t + T_n(t)] = \Phi(u).$$

Разложим $\Phi(u)$ в ряд Тейлора окрестности точки t :

$$\Phi[t + T_n(t)] = \Phi(t) + \Phi'(t) \frac{T_n(t)}{1} + \Phi''(t) \frac{[T_n(t)]^2}{2!} = \Phi(t) + f(t). \quad (3)$$

Поскольку минимальная частота следования импульсов не превышает 30 кГц, можно допустить, что в промежутке между кодовыми импульсами осциллограмма кривой сигнала $f(t)$ изменяется линейно во времени. Это дает возможность пренебречь значениями высших производных в разложении (3).

В первом приближении $\Phi[t + T_n(t)] = \Phi(t) + f(t) T_n(t)$, тогда

$$E = f(t) T_n(t), \quad \text{откуда } T_n(t) = \frac{E}{|f(t)|}. \quad (4)$$

Во втором приближении

$$E = f(t) T_n(t) + f'(t) \frac{T_n^2(t)}{2}. \quad (5)$$

Величина порога срабатывания для гармонического сигнала

$$E_{\min} = \int_0^t A_{\min} \sin \omega t = \frac{A_{\min}}{\omega} (1 - \cos \omega t); \quad E_{\min} = \frac{2A_{\min}(\omega_{\max})}{\omega_{\max}},$$

т. е. минимальный порог определяется порогом слышимости на верхней граничной частоте и этой частотой.

Резюмируя сказанное, запишем функцию преобразования первого приближения:

1) положение во времени n -го импульса $t_n = \sum_1^n T_i$;

2) интервал времени от n до $n + 1$ импульса $T_{n, n+1} = \frac{E}{|f'(t_n)|}$;

3) порог срабатывания связан с амплитудно-частотной характеристикой сигнала соотношением $E = 2A_{\min}(\omega_{\max})/\omega_{\max}$, где A_{\min} — порог слышимости.

Мы получили аналитическое описание дискретного сигнала, который согласно закону Гальбота должен восприниматься человеком как аналоговый сигнал. Эксперименты по проверке этого утверждения проводились на специальном устройстве, реализующем зависимость (4). Блок-схема устройства изображена на рис. 2.

Схема работает следующим образом. Речевой сигнал в аналоговой форме поступает на вход накопительного элемента 4. Накопительный элемент представляет собой интегрирующий усилитель, постоянная времени которого выбирается больше максимального интервала времени между импульсами. Значение сигнала на выходе накопительного элемента пропорционально площади осциллограммы звукового сигнала для каждого момента времени. Выход накопительного элемента подключен на вход триггера Шмидта 5. В исходном состоянии на выходе триггера Шмидта установлен нулевой уровень. При достижении величины входного сигнала, равной порогу срабатывания, триггер Шмидта устанавливается в единичное состояние. Генератор тактовых импульсов 1 формирует серию стандартных прямоугольных импульсов фиксированной частоты, необходимых для формирования двоичного кода. На выходе вентиля 2 будет установлен сигнал высокого уровня только в том случае, когда высокий уровень установится на обоих входах вентиля. Импульс с выхода вентиля подается на вход схемы гашения, которая обеспечивает быстрый разряд накопительного элемента и устанавливает на выходе триггера Шмидта нулевой уровень выходного сигнала. Далее процесс повторяется. Таким образом с выхода вентиля на вход формирователя поступает преобразованный дискретный сигнал. Схема формирователя 3 обеспечивает получение выходных импульсов стандартной длительности.

Как видно из работы схемы, описанная экспериментальная установка — это преобразователь аналог-код интегрирующего типа, который может быть использован для ввода речевого сигнала в ЭВМ. В экспериментах по восприятию дискретного сигнала участвовало три человека. Акустический сигнал, записанный на магнитофон, поступал на преобразователь аналог-код. Дискретный сигнал с выхода преобразователя подавался на телефон.

Экспериментатор имел возможность при помощи специального тумблера предъявлять испытываемому поочередно дискретный и аналоговый сигналы. В первой серии экспериментов на вход преобразователя подавались сигналы синусоидального вида. Осциллографом, подключенным к наушникам, отмечался факт преобразования синусоиды в стандартные импульсы. Уровень аналогового

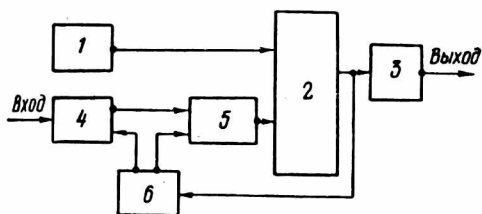


Рис. 2. Функциональная схема преобразователя.

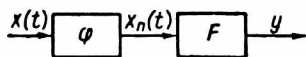


Рис. 3. Блок-схема преобразования слуховой информации.

сигнала равнялся 50 дБ. При этом частота следования импульсов составляла 40 кГц. Во всех опытах было отмечено тождество звучания дискретного и аналогового сигналов, т. е. испытуемые воспринимали дискретный сигнал как чистый тон. Во второй серии экспериментов в качестве входного сигнала использовали дикторскую речь и музыкальные произведения.

Эксперименты показали, что непрерывный и дискретный сигналы по слуховому ощущению неотличимы; это полностью удовлетворяет закону Тальбота.

По своему способу построения дискретный сигнал содержит конечное количество информации. Отсюда можно сделать вывод, что информация, содержащаяся в слуховом ощущении, также конечна и может быть представлена в дискретном виде для ввода в ЭВМ с целью обработки на цифровом уровне. Описанный способ дискретизации позволяет не только ввести информацию в ЭВМ, но также и вывести ее из машины в виде акустического сигнала без предварительного сглаживания. Выведенная таким способом информация может быть воспринята ухом как ощущение, неотличимое от ощущения, формируемого непрерывной осциллограммой.

Таким образом, вывод о дискретности информации, содержащейся в слуховом ощущении, может послужить базой для разработки устройств анализа и синтеза речи. Вторым важным моментом, вытекающим из положения о дискретности слухового

восприятия, является возможность представить преобразованный сигнал в слуховой системе как суперпозицию двух преобразований: 1. Аналого-цифровое преобразование по предложенному нами рецепту. 2. Чисто дискретное преобразование (рис. 3).

Данный вывод не зависит от наших анатомо-физиологических данных, касающихся вопросов импульсной генерации в нейронных сетях, а логически следует из того факта, что дискретный сигнал, формируемый по нашему рецепту, абсолютно не различим на слух от непрерывного сигнала. Обращаясь с этим выводом к техническим задачам, мы получаем возможность вырабатывать рекомендации по моделированию процессов переработки информации слуховым анализатором человека. Согласно схеме (рис. 3) следует сначала производить аналого-цифровое преобразование φ , а уже затем — все дальнейшие преобразования F , требуемые по слуховым моделям. Последние преобразования информации проводятся на дискретном уровне с помощью ЭВМ.

Данные электро-физиологических исследований слухового аппарата позволяют предположить, что такое последовательное соединение преобразователей действительно используется слуховой системой.

Если представление функции в виде суперпозиции двух функций невозможно, значит, невозможна и пробная реализация этих функций в виде последовательного соединения двух блоков. Однако это не исключает возможности приборной реализации двух функций в виде последовательного соединения двух блоков. Выводы структурного характера из функционального результата рискованы и относятся к ним следует осторожно, поскольку они имеют чисто гипотетический характер.

Список литературы: 1. Глушков В. М. Введение в кибернетику. Киев, Изд-во АН УССР, 1968. 323 с. 2. Котельников В. А. О пропускной способности эфира и проволоки в электросвязи.— В кн.: Материалы к 1-му Всесоюзному съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. М., Изд-во Ред. упр. связи РККА, 1933, с. 111—124. 3. Эффект сглаживания в слухе.— Проблемы бионики. Харьков, 1977, вып. 19, с. 31—37. Авт.: А. Я. Абрамов, С. А. Усенко, Ю. П. Шабанов-Кушнарченко. 4. Лабутин В. К., Молчанов А. П. Слух и анализ сигналов. М., 1967, с. 79.