

ТОНАЛЬНЫЙ МЕТОД ОПРЕДЕЛЕНИЯ РАЗБОРЧИВОСТИ РЕЧИ,
ПЕРЕДАВАЕМОЙ ТРАКТАМИ СВЯЗИ

А. Д. Ткаченко

Излагается новый метод определения разборчивости речи, передаваемой трактами связи, без проведения артикуляционных испытаний.

Приводятся теоретическое обоснование метода и необходимые данные для его практического использования; даются результаты сопоставления разборчивости, определенной описываемым методом и артикуляционными испытаниями.

В последнее время вопрос оценки качества речи, передаваемой телефонными системами, приобрел особенное значение. Известно большое разнообразие методов такой оценки, принятых теми или другими организациями; однако среди них особенно важное место занимает метод оценки качества передаваемой речи по разборчивости.

Разборчивость речи, передаваемой телефонными системами, практически определяется методом артикуляции. Суть артикуляционного метода, как известно, состоит в передаче и приеме через испытуемый тракт определенного речевого материала. Разборчивость определяется процентом правильно принятых звуковых сочетаний. В зависимости от характера передаваемого звукового материала может быть определена смысловая разборчивость (или разборчивость фраз) I , словесная разборчивость W , слоговая разборчивость S , звуковая разборчивость D .

Известно также, что между всеми указанными видами разборчивости существует определенная связь, характерная для данного языка. Поэтому для всесторонней оценки испытуемой телефонной системы достаточно произвести определение разборчивости только для одного из указанных видов речевого материала. В практике артикуляционных испытаний наиболее экономичными оказываются методы определения слоговой и звуковой разборчивости; поэтому они получили наибольшее распространение.

Для определения разборчивости методом артикуляции подбирается рабочая группа в составе 4—5 лиц, обладающих необходимым развитием и характеристиками речи и слуха, удовлетворяющими определенным требованиям. Такая группа должна быть предварительно обучена и должна систематически тренироваться в передаче и приеме речевого материала.

Процесс проведения артикуляционных испытаний является длительным. Установлено, например, что для определения слоговой разборчивости, обеспечиваемой испытуемым трактом, указанному составу рабочей группы необходимо принять 1600—2000 слогов [1,2]. Для получения результатов испытаний, т. е. передачи, приема и подсчета правильно принятых слогов, требуется время, в общем случае более 3—4 часов.

Однако несмотря на длительность артикуляционных испытаний, их результаты могут в значительной степени зависеть от субъективных особенностей лиц, входящих в состав группы, а именно, от произношения и поддержания постоянной громкости речи при передаче, напряженности внимания, сообразительности и утомляемости при приеме. Поэтому

полученные непосредственным испытанием результаты нуждаются в обработке, сводящейся к исключению значительно отклоняющихся данных для отдельных пар участников или в полном исключении результатов испытаний отдельных операторов. Это показывает, насколько затруднительно получить сопоставимые результаты испытаний для одних и тех же трактов, проведенных в определенных постоянных условиях, но в разное время и тем более разными рабочими группами. Поэтому артикуляционный метод имеет до сих пор ограниченное практическое применение, несмотря на всеми признанную важность характеристик, получаемых в результате артикуляционных испытаний.

Известен другой, расчетный, путь определения разборчивости, обеспечиваемой телефонными трактами. В работах [2, 3, 4, 5] показано, что разборчивость, обеспечиваемая трактами связи, в принципе может быть определена расчетным путем. Для этого необходимо знать характеристики тракта, измеренные объективными приборами, характеристики шумов и помех, а также характеристики слуха и речи. Однако для реально существующих трактов подобный расчет оказывается сложным и трудоемким, а его результаты вследствие влияния многих изменяющихся факторов лишь приблизительно согласуются с данными практической проверки.

Предлагаемый тональный метод, не требуя проведения трудоемких артикуляционных испытаний, значительно свободнее от влияния субъективных особенностей операторов, принимающих участие в испытаниях, не требует создания рабочих групп и в сравнении с артикуляционным методом занимает в 8—10 раз меньше времени.

Для уяснения сущности предлагаемого метода необходимо рассмотреть основные зависимости, существующие между характеристиками тракта, шумами и помехами, а также характеристиками речи и слуха, с одной стороны, и разборчивостью передаваемой речи — с другой. Метод основывается в этом отношении на известных теоретических и экспериментальных положениях, касающихся восприятия речи слуховым аппаратом человека и распознавания отдельных звуков речи [2, 3, 4, 5, 6].

Известно, что разборчивость обуславливается уровнями интенсивности звуков речи, а также уровнями помех и шумов в пункте приема речи, воздействующих на слух как непосредственно, так и через тракт передачи.

Уровень интенсивности звуков речи зависит от свойств голоса говорящего и электроакустических характеристик тракта; уровень интенсивности шумов и помех зависит от условий приема, а также от свойств тракта. Однако для каждого из двух лиц, ведущих разговор, разборчивость определяется не абсолютной величиной уровня интенсивности звуков речи и шумов, а уровнем ощущения речи, т. е. превышением составляющих речи над порогом слышимости. На разборчивость речи, вообще говоря, влияют и субъективные факторы, отмеченные выше, но при оценке качества трактов они могут быть в той или другой степени исключены путем выбора определенных средних значений этих факторов.

Кроме указанных выше видов разборчивости, вводится понятие формантной разборчивости (A). Формантная разборчивость для данного тракта не может быть определена непосредственным измерением, но может быть вычислена по соответствующим характеристикам речи, шума и слуха. Получив расчетным путем значение формантной разборчивости, можно при помощи известных зависимостей определить любой вид необходимой для практики разборчивости (звуковую, слоговую, словесную, фразовую).

Формантная разборчивость обладает одним свойством, которым не обладает ни один из других видов разборчивости, а именно свойством аддитивности отдельных частных значений, арифметическая сумма которых дает общую разборчивость. Это свойство позволяет найти общую

формантную разборчивость, обеспечиваемую трактом, если известна разборчивость, соответствующая отдельным частотным полосам.

Для каждого национального языка весь частотный диапазон речи может быть разбит на двадцать таких частотных полос, каждая из которых при оптимальных условиях ее передачи обеспечивает 5% формантной разборчивости. Таким образом, передача всего диапазона частот, т. е. двадцати полос, в этих условиях обеспечивает 100% формантной разборчивости A .

Если же уровень ощущения полосы частот (E'_n) не является оптимальным, то формантная разборчивость будет отличаться от 5%, что учитывается коэффициентом восприятия полосы P_n , где $P_n = \varphi(E'_n)$. Формантная разборчивость полосы A с учетом коэффициента восприятия P в этом случае будет $A_n = 0,05P_n$.

Формантная разборчивость всего речевого диапазона A будет равна сумме формантных разборчивостей всех принимаемых полос:

$$A = 0,05 \sum_{n=1}^{n=20} P_n = 0,05 (P_1 + P_2 + P_3 + \dots + P_n).$$

По формантной разборчивости A , как указывалось выше, могут быть определены все другие виды разборчивости на основании зависимостей, установленных для данного языка. Из изложенного видно, что величина формантной разборчивости, а следовательно, и всех других видов разборчивости, определяется уровнем ощущения формант в каждой из 20 частотных полос.

В ряде работ было показано, что формантный спектр с небольшой погрешностью можно выразить через средний (в течение длительного времени) уровень спектра $B_{\text{ср}}$ [3]. Уровень ощущения формант E'_n в каждой полосе может быть тогда определен на основании выражения:

$$E'_n = B_{\text{ср}} - (b + M) - x,$$

где $B_{\text{ср}}$ — средний в течение длительного времени уровень спектра речи на 1 гц, b — затухание тракта, M — фактор маскировки в результате воздействия шумов и помех, x — порог слышимости спектра.

Так как

$$x = \beta_0 - 10 \lg \Delta f_{\text{к}},$$

где β_0 — порог слышимости чистого тона, а $\Delta f_{\text{к}}$ — ширина критической полосы частот, то

$$E'_n = B_{\text{ср}} - (b + M) - \beta_0 + 10 \lg \Delta f_{\text{к}}.$$

Обычно это выражение является основным при расчете разборчивости. При этом $B_{\text{ср}}$, β_0 , $\Delta f_{\text{к}}$ являются величинами известными, b определяется измерением характеристик отдельных элементов тракта или расчетом затухания тракта, M рассчитывается по данным спектра акустических и электрических шумов и характеристикам тракта.

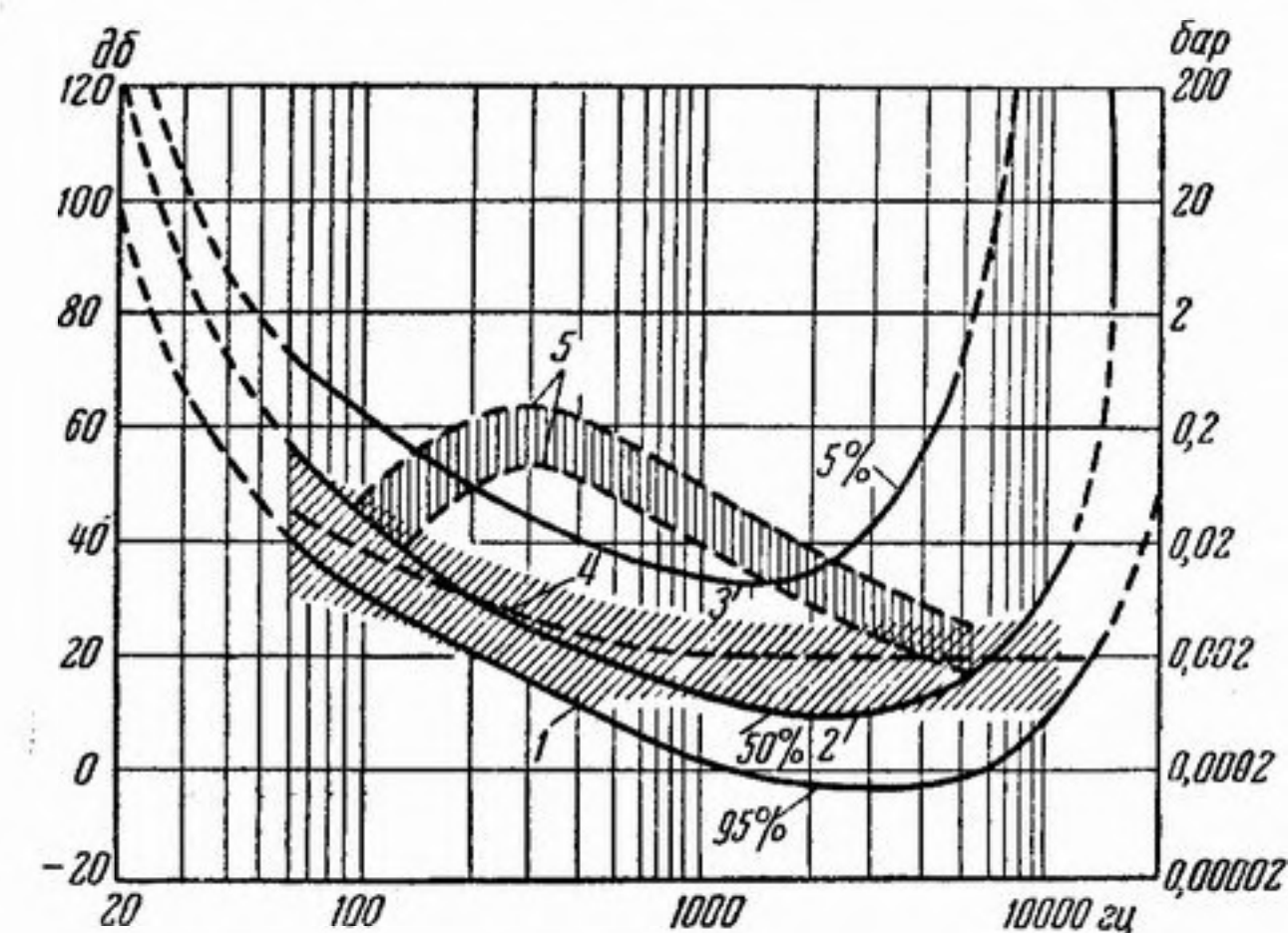
Идея предлагаемого нового метода заключается в том, чтобы определять уровень ощущения речи E'_n (в дальнейшем назовем его превышением) в отдельных частотных полосах путем непосредственного измерения с участием слуха человека.

В принципе задача измерения превышения сводится к определению путем прослушивания на выходе тракта превышений составляющих речи при одновременном воздействии всех мешающих факторов (акустические шумы, электрические помехи и пр.). С этой целью, в заданных условиях акустических шумов и электрических помех, на вход тракта должна подаваться одна из частотных полос спектра речи, а затем уровень в этой

полосе должен доводиться каким-либо способом, например, при помощи магазина затухания, до порога слышимости. Величина затухания $b_{мз}$, установленная в этом случае на магазине затуханий, и будет являться величиной превышения составляющих данной частотной полосы. Подавая на вход тракта последовательно все полосы частотного спектра, можно определить соответствующие им величины превышения.

Проведение указанных измерений можно значительно упростить, если частотные полосы спектра заменить отдельными чистыми тонами, соответствующими по своему положению в спектре речи и по уровню дан-

ным частотным полосам и подаваемыми на вход тракта с помощью искусственного голоса. Из выражения для E'_n видно, что если на данной частоте создавать перед микрофоном чистый тон с уровнем, равным уровню спектра речи $B_{ср}$ для этой частоты, то при измерении величины превышения фактически будут учтены все входящие в это выражение факторы, за исключением логарифмической ширины критической полосы $10 \lg \Delta f_k$.



Фиг. 1. Воздействие шумов на нивелирование порогов слышимости

Поэтому при измерении величины превышения чистого тона необходимо перед микрофоном создавать звуковое давление с уровнем

$$B_{иг} = B_{ср} + 10 \lg \Delta f_k.$$

Для более точного воспроизведения реального спектра речи при помощи искусственного голоса необходимо разбить весь диапазон речи на возможно большее количество частотных полос, приближающихся к критическим, и для средних частот этих полос устанавливать соответствующие величины $B_{иг}$. Практически удобнее иметь в каждой полосе равной разборчивости только одну точку измерений. Поэтому целесообразнее использовать средние частоты полос равной разборчивости, тем более, что ширина этих полос невелика, а современные тракты обладают относительно равномерными частотными характеристиками.

Поскольку при всех измерениях задается определенный спектр речи $B_{иг}$, который должен быть принят единым для всех случаев измерений, то применение искусственного голоса делает данный метод с точки зрения передачи полностью объективным, так как совершенно исключает из передачи участие дикторов.

Работа операторов в пункте приема также значительно облегчается по сравнению с приемом слогов при артикуляционных испытаниях. Задачей операторов в разработанном методе является лишь определение границы слышимости тона. Хотя у разных людей имеется значительное расхождение в порогах слышимости, однако наличие шумов, которые имеют место при эксплуатации, а следовательно, должны существовать и при испытании аппаратуры, приводит к сглаживанию этой разницы. Воздействие шумов на нивелирование порогов слышимости видно из фиг. 1, заимствованной из работы [6]. Кривые 1, 2, 3 являются результатом обобщения измерений порогов слышимости в тишине более чем для полумиллиона человек. При этом цифры 5, 50, 95 означают, что

соответственно 5, 50, 95% человек не слышали тонов, имеющих уровень ниже указанной границы. Кривая 4 дает значения маскирующего уровня за счет наличия комнатных шумов, общий уровень которых составлял 43 дб. Даже для этого случая пороги слышимости для большинства людей фактически перекрываются в большом диапазоне частот (от 200 до 6000 гц). С другой стороны, при испытаниях телефонной аппаратуры устанавливаются, как правило, акустические шумы речевого (сплошного) спектра с уровнем интенсивности не менее 60 дб. Кривые 5 на фиг. 1 показывают значения маскирующего уровня, относящегося к шумам с уровнем 65—75 дб. В этом случае в широком диапазоне частот перекрываются пороги слышимости почти для всех людей, не имеющих явных дефектов слуха.

Нивелирование порогов слышимости разных участников испытаний при наличии шумов приводит к тому, что отпадает необходимость в большом количестве операторов на приеме, так как при определении превышений такие субъективные особенности операторов, как, например, внимательность, быстрота реакции, утомляемость, оказывают существенно меньшее влияние, чем при определении разборчивости путем передачи и приема слогов. Это обеспечивает возможность участия во всех измерениях превышений одного, максимум двух операторов.

Основные исходные данные о характеристиках русской речи были нами заимствованы из работы [1].

Согласно этой работе для получения 100% формантной разборчивости русской речи при приеме ее с оптимальным уровнем ощущения необходимо передать диапазон частот от 120 до 10000 гц. Этот диапазон разделен на 20 частотных полос, каждая из которых обеспечивает одинаковую пятипроцентную формантную разборчивость при оптимальном уровне ощущения E'_n спектра речи в данной полосе.

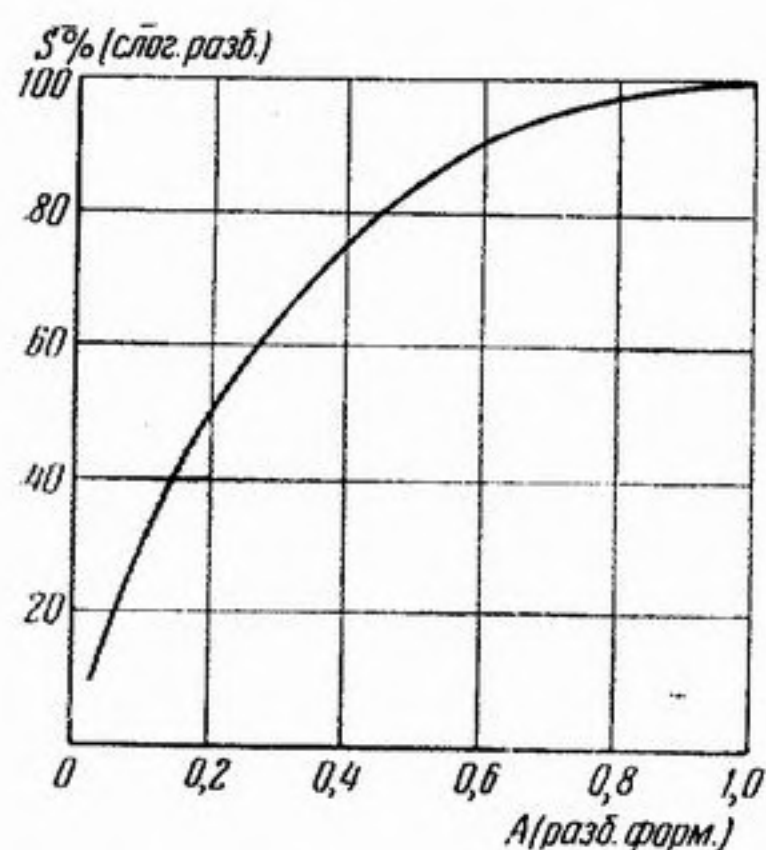
Частотные пределы полос равной разборчивости и их средние частоты приведены в табл. 1. Здесь же приведены значения логарифмической ширины критических полос, относящиеся к средним частотам.

Таблица 1

Частотные полосы равной разборчивости формант

Полосы равной разборчивости, гц	Ширина полос равной разборчивости, гц	Средние частоты полос равной разборчивости, гц	Округленные значения средней частоты, гц	Логарифмическая ширина критической полосы $K = 10 \lg \Delta f_k$, дб
117—510	393	310	300	17,0
510—670	160	590	600	17,2
670—820	150	740	750	17,4
820—975	155	900	900	17,7
975—1130	155	1050	1050	18,1
1130—1290	160	1210	1200	18,4
1290—1445	155	1370	1400	18,6
1445—1600	155	1520	1550	19,0
1600—1760	160	1680	1700	19,3
1760—1925	165	1840	1850	19,6
1925—2100	175	2010	2000	20,0
2100—2300	200	2200	2200	20,4
2300—2540	240	2420	2400	20,7
2540—2820	280	2680	2700	21,0
2820—3160	340	2990	3000	21,5
3160—3530	370	3340	3400	22,1
3530—3930	400	3730	3750	22,8
3930—4400	470	4160	4100	23,0
4400—5400	1000	4900	5000	24,4
5400—10000	4600	7700	8000	—

Зависимость между формантной и слоговой разборчивостью $S = \varphi(A)$ для русской речи, согласно работе [1], приведена на фиг. 2. Зависимость вероятности восприятия от уровня ощущения $P = \varphi[E'_n]$ (фиг. 3) установлена нами по материалам работы [1] с экспериментальным уточнением. Приведенная в указанной работе зависимость $P = \varphi[E'_n]$ являлась лишь первым приближением и поэтому не могла быть непосредственно использована.

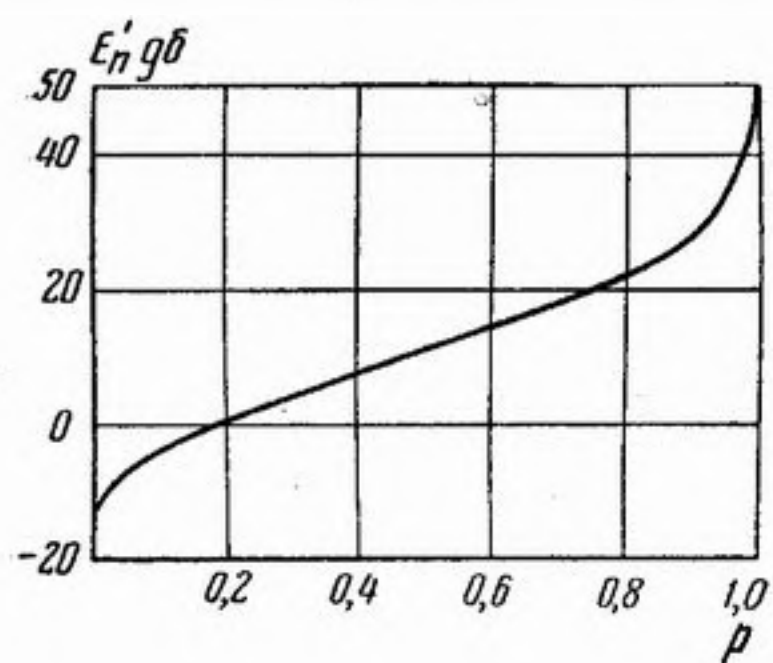


Фиг. 2. Зависимость между формантной и слоговой разборчивостью для русской речи

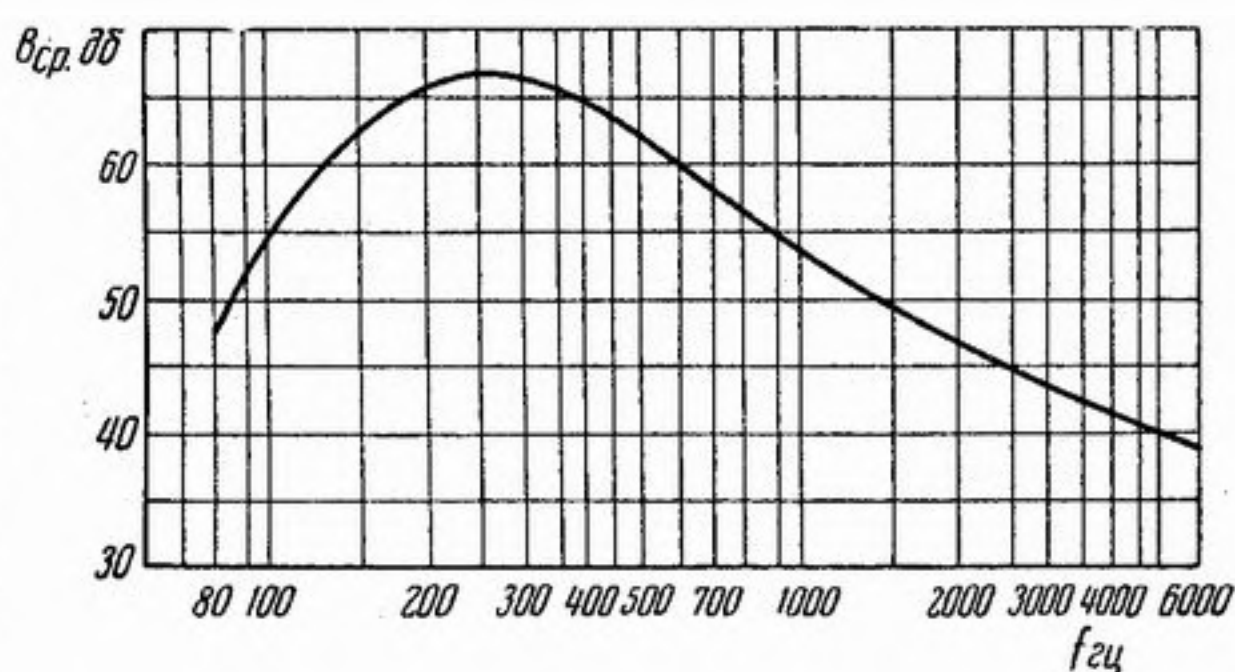
одним тоном, давление, развиваемое искусственным голосом, на расстоянии r должно соответствовать уровню

$$B'_{иг} = B_{ср} + 20 \lg \frac{R}{r} + 10 \lg \Delta f_k.$$

В табл. 2 приведены значения давлений, которые должен развивать искусственный голос для воспроизведения спектра русской речи $B'_{ср}$ на



Фиг. 3. Зависимость вероятности восприятия от уровня ощущения



Фиг. 4. Частотный спектр русской речи

расстоянии 2 см от рта. В графе 6 табл. 2 давления изменены в соответствии с изменением $B'_{ср}$ на 12 дБ. Это сделано с целью упрощения работы операторов по определению превышения, о чем будет сказано ниже. Фактически этим учитывается значение пикфактора речи и получение не средних, а пиковых значений уровней речи, вследствие чего по ординатам графика $P = \varphi(E'_n)$ фиг. 3 должны быть сделаны соответствующие поправки на 12 дБ, т. е., например, уровень — 12 дБ на фиг. 3 должен соответствовать теперь 0 дБ.

Перейдем теперь к рассмотрению практической стороны использования предлагаемого метода. Для измерения составляющих речи над по-

Таблица 2

Значение давлений, эквивалентных давлениям в спектре речи, развиваемых искусственным голосом

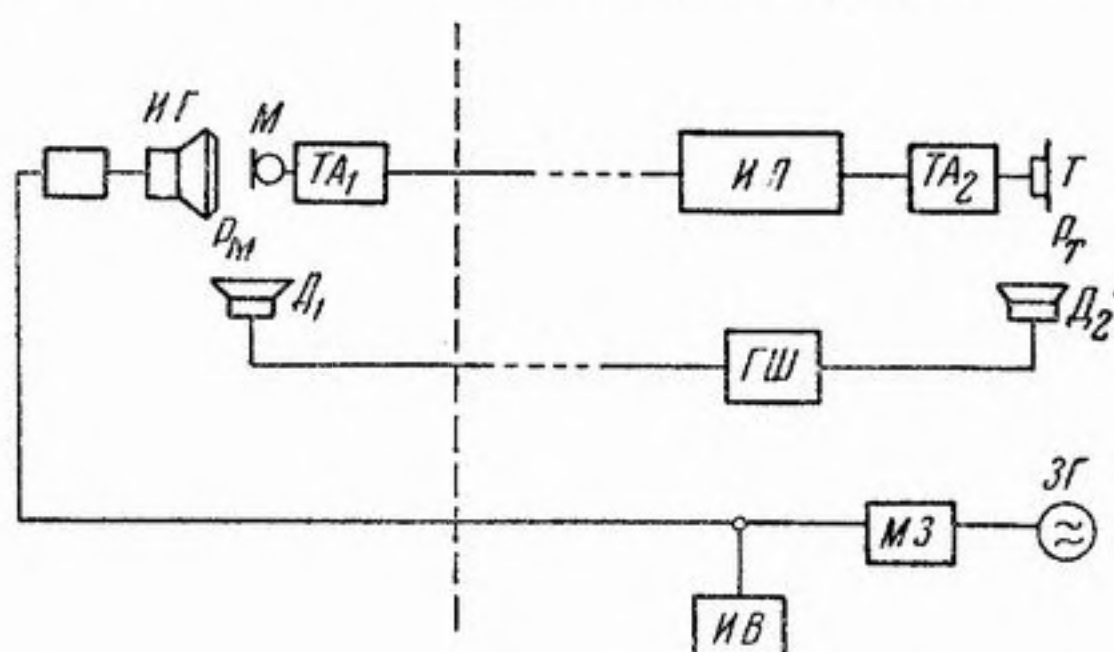
Средняя частота, гц	Средний уровень спектра речи (на расстоянии 8 см), дб	Логарифмическая ширина критической полосы $K=10 \lg \Delta f/k$, дб	Уровень интенсивности искусственного голоса на расстоянии 2 см, дб	Давление на расстоянии 2 см, бар	Давление на расстоянии 2 см, увелич на 12 дб, бар
300	67	17,0	96	12,5	50
600	61,5	17,2	90,6	6,8	27,2
750	58	17,4	87,5	4,7	18,8
900	56	17,7	85,4	3,7	14,8
1050	55	18,1	85,0	3,5	14
1200	52	18,4	82,6	2,7	10,8
1400	51	18,6	81,6	2,4	9,6
1550	50	19,0	81,0	2,2	8,8
1700	48	19,3	79,0	1,8	7,2
1850	47,8	19,6	79,2	1,8	7,2
2000	47	20,0	79,1	1,8	7,2
2200	46	20,4	78,7	1,7	6,8
2400	45,5	20,7	78,2	1,6	6,4
2700	44	21,0	77,6	1,5	6,0
3000	43,5	21,5	77,2	1,4	5,6
3400	42,5	22,1	76,6	1,35	5,4
3750	42,0	22,8	76,8	1,4	5,6
4100	41,5	23,0	76,7	1,4	5,6
5000	40,0	24,4	76,4	1,3	5,2

рогом слышимости, т. е. для определения уровня ощущения, используется схема, изображенная на фиг. 5.

Требуемые звуковые давления P_M согласно табл. 2 на соответствующих частотах создаются искусственным голосом ИГ у микрофона испытуемого тракта M .

Фиг. 5. Схема к определению уровня ощущения

ИГ — искусственный голос, M — микрофон испытуемого аппарата, TA_1 — передающий телефонный аппарат, ИЛ — искусственная линия, T — телефон испытуемого аппарата, TA_2 — приемный телефонный аппарат, ГШ — генератор шумов, D_1 и D_2 — динамики для воспроизведения акустических шумов, ЗГ — звуковой генератор, МЗ — магазин затухания, ИВ — измерительный вольтметр



пытуемого тракта M . На выходе тракта (телефон T) прослушивается звук данного тона. При прослушивании тона оператор при помощи магазина затухания $MЗ$, включенного в цепь искусственного голоса, устанавливает пороговое значение тона и по показаниям магазина затухания определяет превышение.

Как видно из схемы фиг. 5, для измерения превышений необходимо иметь два отдельных звукоизолированных помещения, а также обеспечить необходимые шумы, характеристики и уровни которых соответствуют требованиям.

Все приборы размещаются так, чтобы один оператор мог проводить измерения без посторонней помощи. Искусственный голос должен быть предварительно отградуирован, чтобы можно было устанавливать давления $P_{иг}$ по напряжению, подаваемому на искусственный голос и измеряемому при помощи вольтметра ИВ. При этом целесообразно скоррек-

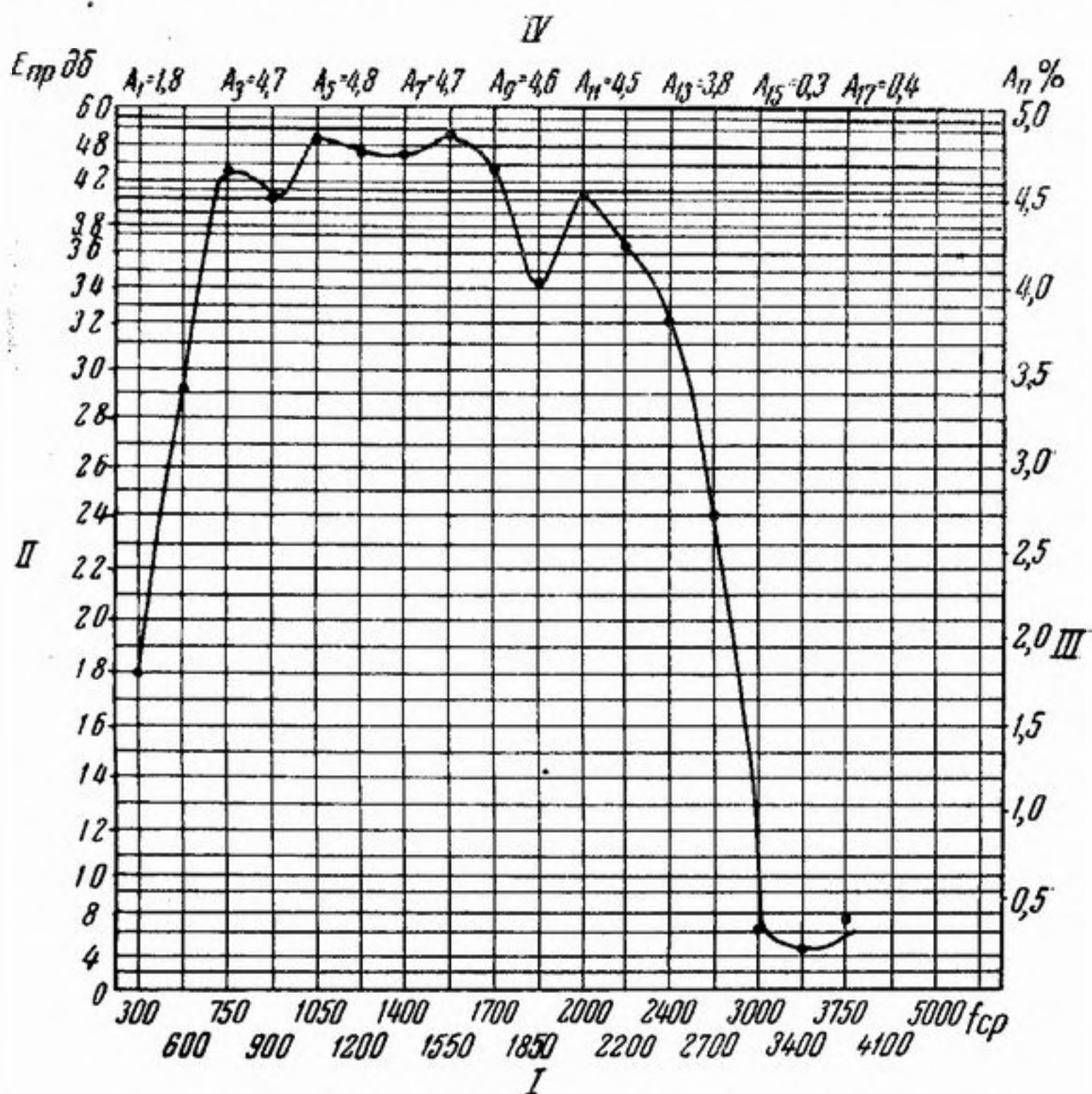
тировать характеристики искусственного голоса так, чтобы при постоянном напряжении получались требуемые значения $P_{иг}$.

Измерив превышения составляющих речи E'_n для средних частот полос равной разборчивости и определив для каждой из них согласно графику фиг. 3 коэффициент восприятия P , вычисляют формантную разборчивость:

$$A = 0,05 \sum_{n=1}^{n=20} P,$$

а по графику $S = \varphi(A)$ (фиг. 2) находят слоговую разборчивость S .

Для упрощения определения формантной разборчивости более целесообразно пользоваться разработанным для этой цели бланком, показан-



Фиг. 6. Бланк для упрощения определения формантной разборчивости речи

ным на фиг. 6. По оси абсцисс снизу отложены средние частоты полос равной разборчивости, по оси ординат слева — уровни ощущения (превышения) E'_n ($E_{пр}$) и справа — значения формантной разборчивости для полосы A_n , соответствующие определенному значению E'_n . Сверху вписываются значения формантной разборчивости данной полосы для найденного значения E'_n . При пользовании таким бланком, измерив превышение, наносят его на бланк против соответствующей частоты и уровня E'_n . После определения всей характеристики по измеренным превышениям определяется формантная разборчивость для каждой полосы и записывается ее значение против соответствующей частоты.

Просуммировав значения формантной разборчивости для всех частотных полос A_n , оператор определяет общую формантную разборчивость. Далее, как было указано выше, на основании графика фиг. 2 определяется слоговая разборчивость, которая записывается на тот же бланк.

В качестве примера на бланк фиг. 6 нанесена одна из характеристик превышений и определена для нее слоговая разборчивость. Для данного случая $\Sigma A_n = 58,3$ а $S = 89\%$. Из этого примера видно, что характеристика превышений имеет и самостоятельный интерес, так как она показывает действительное ограничение полосы передаваемых частот.

Указанная схема измерения превышений, когда пороговые значения сигналов на выходе определяются пороговым значением сигнала на входе тракта, учитывает также возможную нелинейность амплитудной характеристики тракта.

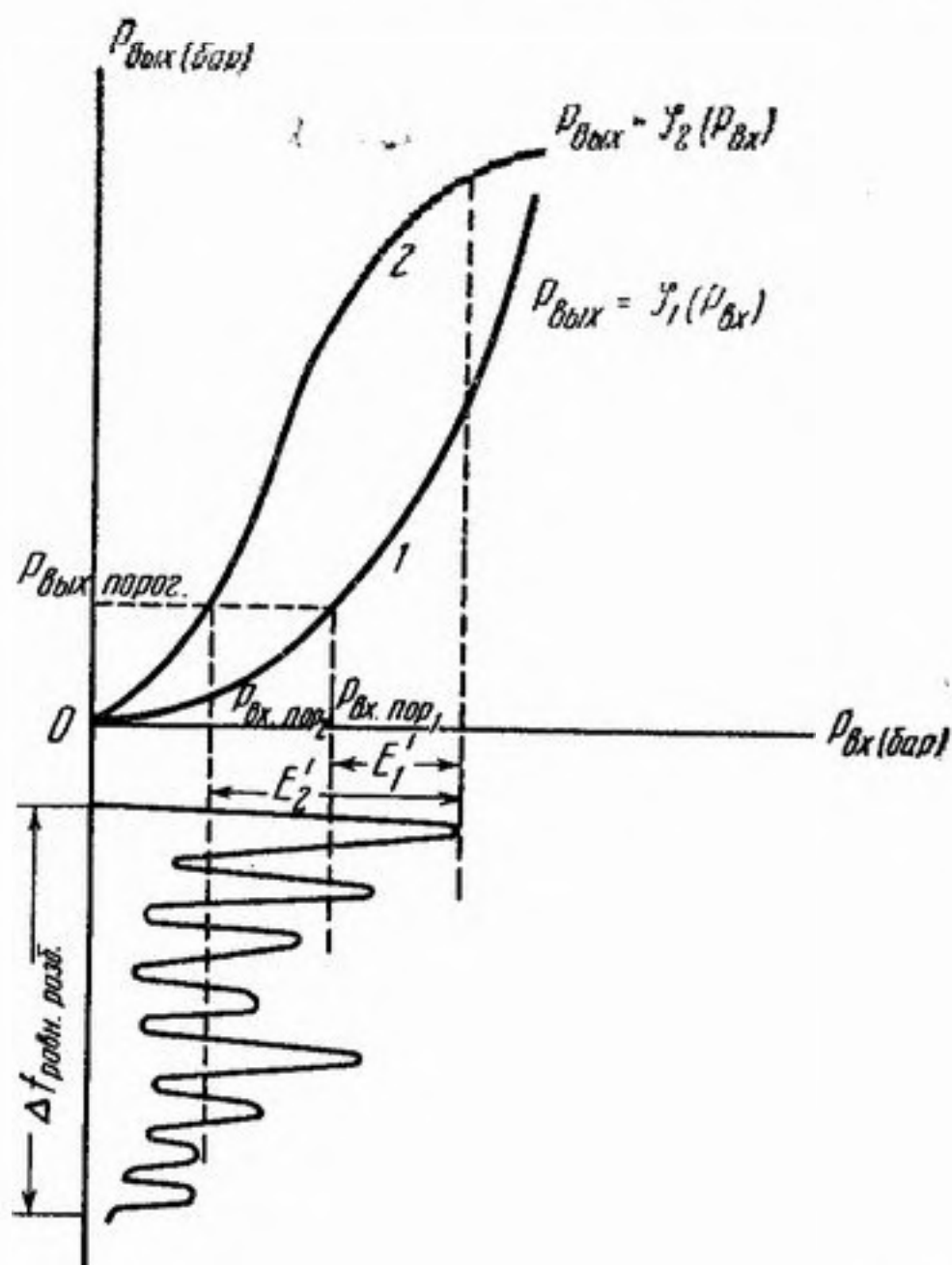
Измерением превышения фактически определяется вероятность восприятия формант, т. е. относительное количество слышимых формант.

На фиг. 7 даны две произвольные амплитудные характеристики тракта. Если пороговое значение звукового давления на выходе (в телефоне) при данных условиях приема будет $P_{\text{вых. пор}}$, то пороговое значение звукового давления на входе для амплитудной характеристики 1 будет $P_{\text{вх. пор}_1}$, а для амплитудной характеристики 2 будет $P_{\text{вх. пор}_2}$. Ясно, что тракт с амплитудной характеристикой, подобной характеристике 2, обеспечивает прохождение звуковых сигналов, отвечающих меньшему звуковому давлению, чем тракт с характеристикой 1.

С целью практической проверки универсальности разработанного метода были проведены эксперименты на специальном тракте, а также на трактах, составленных из реальных телефонных аппаратов. Экспериментальный тракт с электродинамическим микрофоном и телефоном позволял при помощи фильтров, дискриминатора и магазина затуханий изменять в широких пределах частотную полосу и уровень передачи. При проведении экспериментов в пункте приема создавались акустические шумы речевого спектра с уровнем интенсивности от 65 до 90 дб. Это обеспечивало изменение слоговой разборчивости речи от 8 до 98%. Для каждого варианта частотной характеристики снимались указанным методом превышения и проводились артикуляционные измерения группой операторов.

При проведении экспериментов с реальными телефонными аппаратами, с угольным микрофоном, измерения производились при различном затухании искусственной линии и шумах с уровнем интенсивности 65 ÷ 90 дб. В целях более точного сопоставления измерялся средний уровень спектра речи $V_{\text{ср}}$ диктора (мужчины), участвующего в артикуляционных испытаниях. Спектр речи данного диктора принимался в качестве исходного при измерениях превышений.

Результаты измерений разборчивости артикуляционным методом и по превышениям приведены в табл. 3. Из сопоставления результатов видно, что тональный метод может с успехом заменить артикуляционные испытания телефонной аппаратуры в лабораторных условиях и при заводской приемке. На измерение затрачивается не больше 30—40 мин., причем измерение производится одним оператором.



Фиг. 7. К сравнению двух трактов с различными амплитудными характеристиками

Таблица 3

Результаты сопоставления разборчивости, полученной двумя методами

Для экспериментального тракта

Слоговая разборчивость по данным артикуляционных испытаний в %	8	17	27,5	26,4	33	48	58	60	77	86	97
Слоговая разборчивость по измерению превышений в %	10	20	31	27	36	46,5	57	64	73	85	98

Для реальных телефонных аппаратов

Слоговая разборчивость по данным артикуляционных испытаний в %	28,5	43	49	55,3	58	73,5	85
Слоговая разборчивость по измерению превышения в %	32	42	51	58	60	77,5	89

Простота измерений и измерительной аппаратуры делают данный метод удобным для проведения испытаний аппаратуры связи по разборчивости передаваемой речи.

ЛИТЕРАТУРА

1. Труды ВКАС, 1951, № 29—30.
2. Расчет и измерение разборчивости русской речи, Труды ВКАС, 1952, № 33.
3. В. К. Иофе, М. А. Сапожков. К вопросу о методике расчета разборчивости речи. Труды Комиссии по акустике, 1951, № 6.
4. Ю. С. Быков. Труды ЦНИИС СА, 1948, № 1.
5. N. French and I. Steinberg. Факторы, определяющие разборчивость речи, J. Acoust. soc. Am., 1947, 19, 91.
6. H. Fletcher, Слышимость как фактор, определяющий точность воспроизведения, звуков. Proc. Inst. Radio Eng., 1942, 30, 266.

Москва

Поступила в редакцию
14 апреля 1952 г.