архив журнала "Звукорежиссер" : 2002 : №;5

Часть 17.4

Слух и речь, ч 4.
Субъективные и объективные методы оценки разборчивости речи

Ирина Алдошина

Как уже было отмечено в предыдущих статьях "Слух и речь", речевой сигнал имеет двойственную структуру: с одной стороны - это обычный акустический сигнал, объективные акустические параметры которого вызывают определенные субъективные ощущения. Взаимодействия между ними, в соответствии с общими психофизическими законами, неоднозначны и нелинейны.

С другой стороны, речевой сигнал имеет особую структуру, в которой закодирована семантическая (смысловая) информация. Поэтому процесс слухового восприятия речи представляет собой, прежде всего, процесс расшифровки и распознавания семантического и эмоционального содержания информации, содержащейся в речевом сигнале. Исследование этого процесса, то есть того, как мозг переводит акустические признаки речевого сигнала в его фонетическое и смысловое содержание, является в настоящее время одной из самых актуальных проблем в современной науке.

Современные достижения в цифровой обработке сигналов позволили достичь значительных успехов в этой области, и получить практические результаты в компьютерном распознавании и синтезе речевых сигналов. Понимание процессов слухового восприятия речевого сигнала и расшифровки его смыслового содержания являются чрезвычайно важными для практики работы звукорежиссеров, поскольку в процессе работы с речью и пением необходимо понимание того, какие признаки в них являются наиболее критичными для передачи смыслового содержания. Однако, поскольку современные технологии расшифровки и синтеза речевых сигналов появились только в последние годы и еще достаточно сложны для применения, на протяжении уже нескольких десятилетий используются интегральные методы оценки правильной передачи смысловой информации, заключенной в речевом сигнале (в т.ч. и в вокальной речи - пении) , - это *методы оценки разборчивости*. Поэтому остановимся на субъективных и объективных методах оценки разборчивости речи, а затем уже обратимся к расшифровке спектрограмм и современным теориям восприятия речи.

Оценка разборчивости необходима при разработке и использовании различных систем звукоусиления, при оценке акустического качества помещений (театральных и концертных залов, студий, кинозалов и др.) , поскольку, в конечном итоге, качество зала определяется тем, насколько слушатели хорошо понимают смысловое содержание речи, пения и музыки. Разумеется, понимание смыслового содержания не исчерпывает всех аспектов восприятия речи - в ряде случаев не менее важным является передача ее эмоционального содержания (тембра, интонации, темпа и др.). Вопрос о связи акустических характеристик речи (особенно пения) с ее эмоциональным содержанием является чрезвычайно интересной проблемой, и о ней будет рассказано в дальнейшем.

Не менее важна оценка разборчивости и для построения различных коммуникационных систем (радиовещательных, телефонных и др.) . Как показывает опыт работы звукорежиссеров, вопросы, как обеспечить хорошую разборчивость в различных залах, особенно в тех, где установлены системы звукоусиления, являются чрезвычайно актуальными.

В соответствии с международными стандартами, в частности ISO/TR 4870, под *разборчивостью* понимается *"степень, с которой речь может быть понята (расшифрована) слушателями"*. Под этим понимается степень, с которой слушатели могут идентифицировать (понять смысл) фраз, слов, слогов и фонем. В соответствии с этим различаются виды разборчивости: фонемная, слоговая, словесная и фразовая, которые, однако, все связаны друг с другом, и могут быть пересчитаны одна в другую.

При передаче речевого сигнала происходит неизбежная потеря информации. Хотя речевой сигнал обладает определенной избыточностью, однако различные шумы, искажения и реверберационные помехи могут привести к настолько значительной потере информации, что это сделает невозможным понимание смысла речи. Следует отметить, что "слышимость" и "разборчивость речи" - это разные понятия. Речь может звучать очень громко и быть прекрасно слышна, но быть при этом совершенно неразборчивой (например, в залах вокзалов, аэропортов и др.). Поэтому для оценки разборчивости речи разрабатываются специальные методы, отличные от оценок ее громкости, и разработкой этих методов занимаются крупные международные организации: ISO, AES, IEC и др.

Все известные в настоящее время методы оценок разборчивости могут быть разделены на две большие группы: *субъективные экспертные методы* (ГОСТ 25902-83, ГОСТ 51061-97, стандарт ANSI S3.2 и др.), и *объективные методы*, основные из которых: **%Alcons** - процент артикуляционных потерь со- гласных (percentage Articulation Loss of Consonants); **AI** - индекс артикуляции (articulation Index); **STI** - индекс передачи речи (speech transmission index); **RASTI** - быстрый индекс передачи речи (rapid speech transmission index); **SII** - индекс разборчивости речи (speech intelligibility index) и др. (стандарты ISO/TR-4870, ANSI S3.2, S3.5; IEC 268-16 и др.) .

Остановимся на том, какие основные факторы влияют на уровень разборчивости речи в различных системах коммуникации и звукоусиления.

Средние интегральные характеристики речи показаны на рисунке 1. Из них видно, что основная энергия речи сосредоточена в полосе до 2 кГц. График распределения амплитудного состава речи показывает, что более 80% звуков речи имеют уровень меньше 50 дБ, и легко могут маскироваться шумами. Среди этих звуков могут оказаться согласные звуки - самые информативные. Гласные звуки имеют основную частоту фонации в пределах 80: 250 Гц, и значительная часть их энергии сосредоточена в формантных областях в пределах 450:4000 Гц. Именно по распределению формантных областей в спектре и происходит распознавание гласных звуков. Хотя гласные звуки имеют длительность 30:300 мс, и именно в них сосредоточена основная энергия речевого сигнала, основной вклад в разборчивость вносят согласные звуки, которые имеют значительно меньшую длительность, от 10 до 100 мс. Они ниже по уровню на 27 дБ и их спектр - особенно у шумовых (С, З) и взрывных (Д, Т) согласных - расположен в основном в высокочастотной области 2:10 кГц. Ключевую роль в распознавании речи играют октавные полосы в области 1, 2, 4 кГц. Они содержат до 75% речевой информации. Особо важную роль играет октавная полоса в области 2 кГц - до 33% речевой информации. Следует отметить также, что реальные речевые источники (например, диктор) имеют характеристику направленности в пределах угла покрытия 120о в горизонтальной и 90о в вертикальной плоскостях с коэффициентом направленности Q = 2,5 в области 2 кГц. Это имеет существенное значение для разборчивости.

Среди многочисленных факторов, влияющих на разборчивость речи, прежде всего можно выделить следующие:

1. *Маскирование* другими звуками, в том числе шумами в реверберирующем помещении и др. Шумы могут создаваться вентиляцией, внешними проникновениями, шумами аппаратуры, публикой, электронной аппаратурой и др.

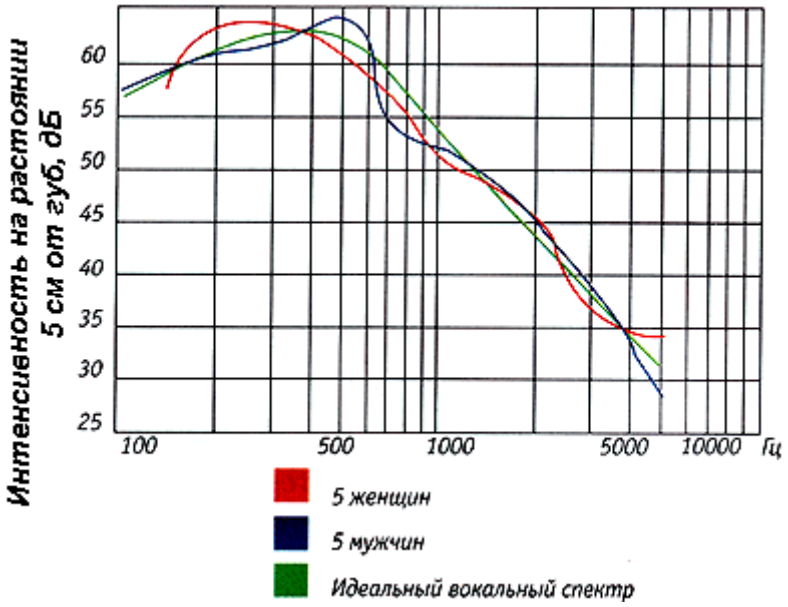

Рис. 1 Частотная зависимость спектральной плотности речи

Процент потери разборчивости зависит, прежде всего, от отношения уровня речевого сигнала к уровню шума (S/N), которое должно быть выше определенного уровня, чтобы можно было понять смысловое содержание речи. Степень маскировки шумом будет зависеть от отношения S/N и от спектрального состава шума. Для широкополосного шума (20:4000 Гц) зависимость процента словесной разборчивости от S/N показана на рисунке 2. Из него видно, что процент словесной разборчивости будет больше 80% только при отношении S/N > 12 дБ.

Если шум узкополосный, то степень маскирования речи и потеря разборчивости зависят от частотной полосы (рисунок 3), то более "опасными", чем высокочастотные (1800:2500 Гц) шумы, являются низкочастотные шумы (135:400 Гц).

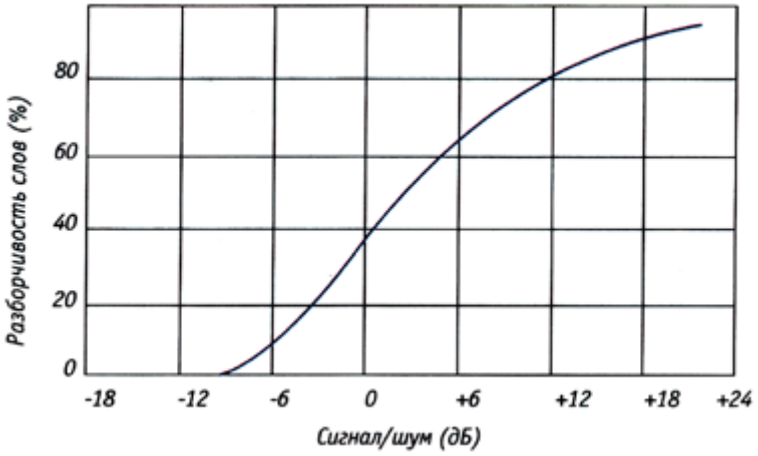

Рис. 2 Зависимость словесной разборчивости от отношения сигнал/шум для широкополосного маскирующего шума

Сильное воздействие на разборчивость речи оказывает шум от других голосов (шум толпы) (рисунок 4). Поскольку этот шум сходен с речью по спектральному составу, то, как следует из графика, уровень словесной разборчивости резко снижается, особенно при увеличении числа мешающих голосов. Именно поэтому "эффект близости" (proximity effect) у направленных микрофонов, связанный с увеличением чувствительности на низких частотах при приближении микрофона к источнику звука (попадание в зону сферической волны), приводит к значительной потере разборчивости за счет маскировки низкочастотными составляющими речевого сигнала. Поэтому необходимо применение высокочастотных фильтров с крутизной 12дБ/окт и с частотой среза не ниже100 Гц.

Влияние шумов на разборчивость речи зависит также от направления их прихода: если направления речевого сигнала и шума совпадают, то степень маскировки и, соответственно, процент потери разборчивости будет наибольшим. Слуховой системе трудно провести их разделение, но чем больше расстояние между ними, тем выше разборчивость.

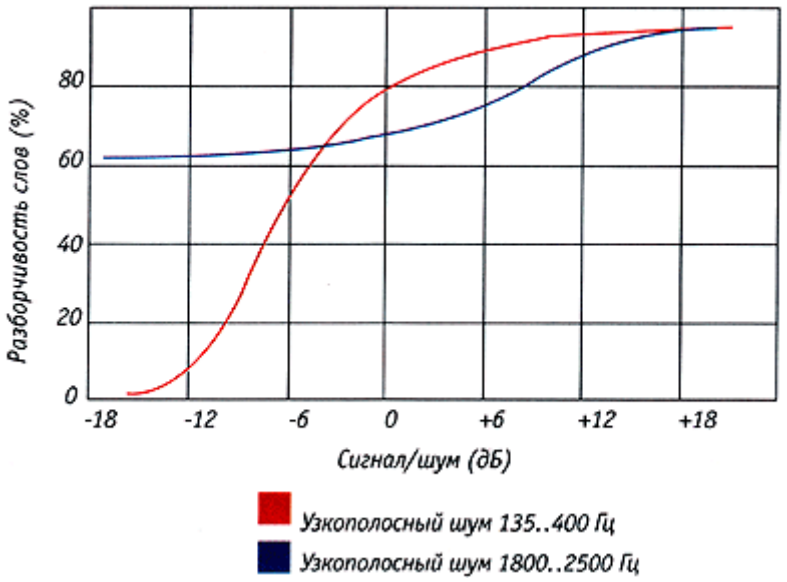

Рис. 3 Зависимость словесной разборчивости от отношения сигнал/шум для н/ч и в/ч узкополосного шума

2. *Процесс реверберации* в помещении оказывается критическим для разборчивости речи, поскольку в ту же точку, где расположен слушатель, приходят со всех сторон отраженные сигналы с похожей спектральной структурой, но с большим содержанием низкочастотных составляющих. Особенно это заметно в тех местах помещения, где расстояние дальше критического "радиуса гулкости", на котором энергия прямого сигнала равна энергии отраженных сигналов. Как известно, для каждого вида музыки и речи имеется свое оптимальное время реверберации (время, в течение которого уровень сигнала спадает на 60 дБ). Примеры для некоторых видов музыки и речи в помещениях различных объемов показаны на рисунке 5. Как видно из графика, оптимальное время реверберации для речи существенно ниже, чем для музыки, и находится в пределах 0,4:0,8 с. Прослушивание речевых сообщений в помещениях с большой реверберацией приводит к значительной потере разборчивости (например, в залах вокзалов, соборах).

Существенную роль для повышения разборчивости играет отношение прямого звука к реверберирующему звуку на всей площади слушательских мест: чем выше уровень прямого звука по отношению к уровню реверберирующего звука, тем выше процент разборчивости. Отсюда вытекают особые требования к выбору характеристик направленности систем звукоусиления. Кроме того, следует отметить, что существенную роль играет также отношение энергии ранних отражений (прибывающих к слушателю в первые 80:100 мс), к энергии поздних отражений - именно поэтому рекомендуется установка дополнительных отражающих экранов у трибуны оратора и у сцены драматических театров.

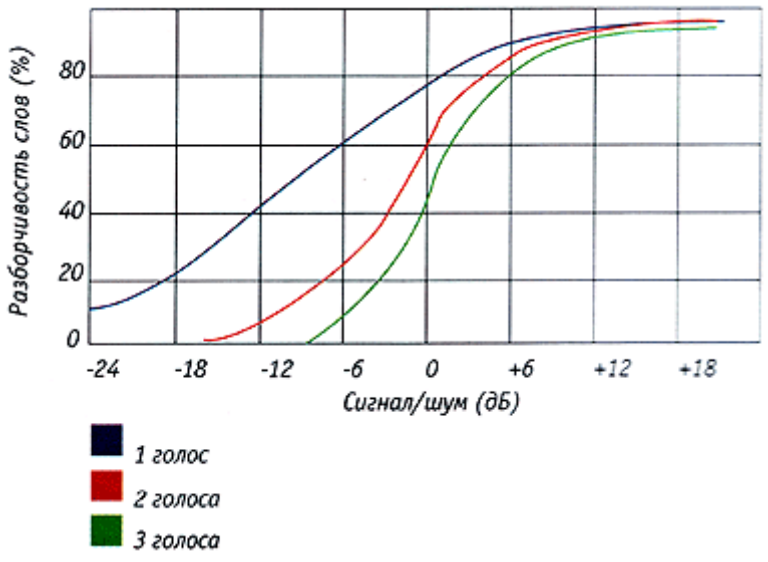

Рис. 4 Зависимость словесной разборчивости от отношения сигнал/шум при воздействии других голосов

3. *Параметры тракта звукоусиления*, такие, как частотный диапазон, форма частотной характеристики тракта, уровень нелинейных искажений, фазовые искажения и др., имеют существенное значение для обеспечения хорошей разборчивости речи. Для высококачественной передачи речи необходимо обеспечить частотный диапазон от 80 Гц (час-тота фонации низких мужских голосов) до 10 кГц (спектры шумовых согласных). Разумеется, определенный процент разборчивости сохраняется и при ограничении полосы пропускания, например, в полосе от 300 Гц до 3 кГц (используется в телефонной связи), хотя становятся трудно различимыми согласные звуки "т" и "д", "с" и "ф", и др.

Ниже 80 Гц АЧХ должна быть резко ограничена для уменьшения уровня маскировки . В пределах указанной полосы АЧХ должна быть плоской (для музыки в некоторых системах звукоусиления делается спад к высоким частотам), но для речи это уменьшает спектральный уровень согласных, который и так мал. Кроме того, должна быть малой неравномерность АЧХ, поскольку значительные пики и провалы могут привести к потере наиболее ценной информации в диапазоне формантных областей гласных, или в области максимальной энергии согласных звуков. Выполненные за последнее время исследования показали достаточное влияние фазовых характеристик тракта на разборчивость речевого сигнала, также как и на восприятие тембра и высоты тона. Поэтому требования к линейности фазовых характеристик тракта также являются существенными.

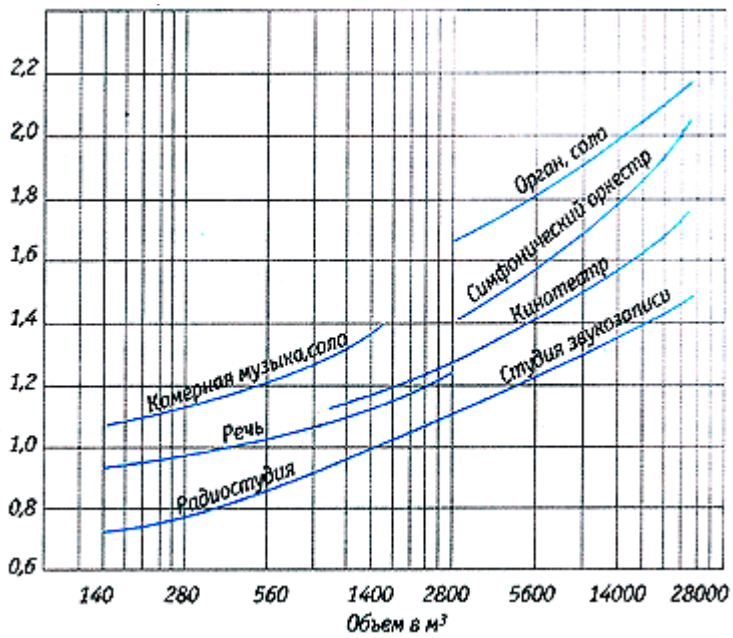

Рис. 5 Оптимальное время реверберации для помещений разного объема

Различные виды нелинейных искажений при обработке сигнала в системах звукоусиления могут значительно снизить разборчивость речи: например, влияние клиппирования на процент словесной разборчивости речи показан на рисунке 6. При этом появляются дополнительные гармоники, которые маскируют речь. Наиболее существенное влияние на разборчивость оказывают интермодуляционные искажения в системе, так как возникают суммарные и разностные тоны, негармонические к основному тону, что существенно маскирует речевой сигнал.

Таким образом, на разборчивость речи в различных помещениях влияют следующие основные факторы: отношение сигнал/шум, время реверберации, уровень прямого звука, отношение энергии ранних и поздних отражений, частотный диапазон системы звукоусиления, формы АЧХ и ФЧХ, характеристики направленности, уровень нелинейных (особенно интермодуляционных) искажений, равномерность покрытия площади прослушивания.

Для количественной оценки разборчивости речи применяются как субъективные методы (экспертные оценки), так и объективные (расчет целого ряда параметров). Хотя за последние годы введено достаточно много новых объективных критериев и созданы специальные компьютерные программы для их реализации, оценки разборчивости речи с помощью квалифицированных экспертов по прежнему остаются наиболее достоверными, и все новые объективные критерии сравниваются с ними.

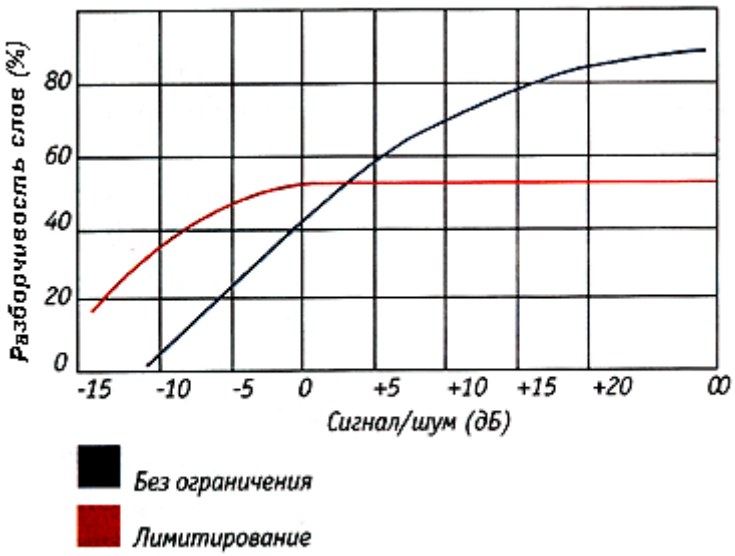

Рис. 6 Влияние клиппирования сигнала на процент словесной разборчивости речи

Субъективные методы оценки разборчивости

Оценка процента разборчивости (артикуляции) зависит от ряда факторов, основные из которых следующие:
- выбор для прослушивания элементов речи (звуки, слоги, слова, фразы), наиболее полно отражающих статистику данного языка;
- подбор состава экспертов и степень их тренированности;
- качество голоса диктора, его дикция, интонация и др.;
- требования к помещению и условиям в нем (уровню шумов и др.)
- методика проведения измерений и методы статис-тической обработки результатов.

Именно эти требования и задаются в различных стандартах, так как только при их точном соблюдении можно получить повторяемость результатов.

Для регламентации таких испытаний введены отечественные стандарты: ГОСТ 25902-83. "Зрительные залы. Методы определения разборчивости речи", ГОСТ51061-97 "Параметры качества речи и методы ее измерения", международные стандарты ISO/TR4870, IEC 268-16. Сейчас разрабатывается новый стандарт AES, а также многочисленные национальные стандарты, например американский стандарт ANSI S3.2-1989 - "Method for measurement the Intelligibility of Speech Over Communication Systems" (имеется новая редакция R-1999) .

Стандартизованные правила прежде всего касаются *отбора испытательного материала*: специально составленных таблиц фраз, слов или слогов, которые записываются или передаются диктором для оценки помещения, системы звукоусиления, или других систем коммуникации. В зависимости от типа используемых при испытаниях элементов речи различается разборчивость *звуковая, слоговая, словесная* и *фразовая*. Все эти виды разборчивости при испытании одной и той же системы будут выражаться разными числовыми величинами, так как процент правильных оценок для предсказуемого сообщения всегда выше, чем для непредсказуемого. Степень предсказуемости при прослушивании фразы выше, чем при слушании отдельных слов или слогов, поскольку если часть фразы не услышана, то можно догадаться по смыслу о ее содержании. В связи с этим находятся и соотношения соответствующих видов разборчивости: фразовая - выше словесной, словесная - выше слоговой, слоговая - выше фонемной.

На рисунке 7а показана зависимость фразовой разборчивости от словесной, на рисунке 7б -

словесной от слоговой. Из-за наличия таких связей для оценки разборчивости можно использовать различные элементы речи, однако в отечественных стандартах чаще используется оценка слоговой разборчивости, поскольку она имеет ряд преимуществ (меньшую запоминаемость, удобство при обработке и др.).

При проведении таких испытаний специально подобранные дикторы (с хорошей дикцией, правильной речью, с хорошим слухом) зачитывают в определенном ритме стандартизованные слоговые таблицы в выбранном помещении - с естественной акустикой или через звукоусилительную систему. Желательно, чтобы эксперты были незнакомы с дикторами, так как. разборчивость у знакомых дикторов выше за счет запоминания экспертами их интонации, дикции и др. Количество дикторов должно быть не менее четырех, причем желательно, чтобы они имели минимальную разницу по акустическим характеристикам голосов. Для проведения испытаний группа слушателей размещается в разных местах помещения и записывает прослушиваемый текст. Отношение правильно записанных на слух фонетических элементов к общему количеству переданных и определяет процент разборчивости.

Для получения статистически достоверных результатов необходимо привлечение достаточно большого числа слушателей. В стандарте ГОСТ 25902-83 принята численность группы слушателей в 20 человек, позволяющая получить статистически надежные результаты. Для зала вместимостью более двух тысяч человек привлекаются две группы слушателей, а если вместимость зала более пяти тысяч человек - три группы слушателей, по 20 человек в каждой группе. Для сокращения времени испытаний в каждой группе проводится цикличная смена мест, при которой каждый слушатель с занимаемого им места переходит на место другого эксперта. Цикл заканчивается, когда все слушатели побывают на всех местах испытаний. Места, на которых определяется разборчивость, должны быть равномерно распределены по залу, а их количество должно соответствовать числу участвующих в испытаниях слушателей.

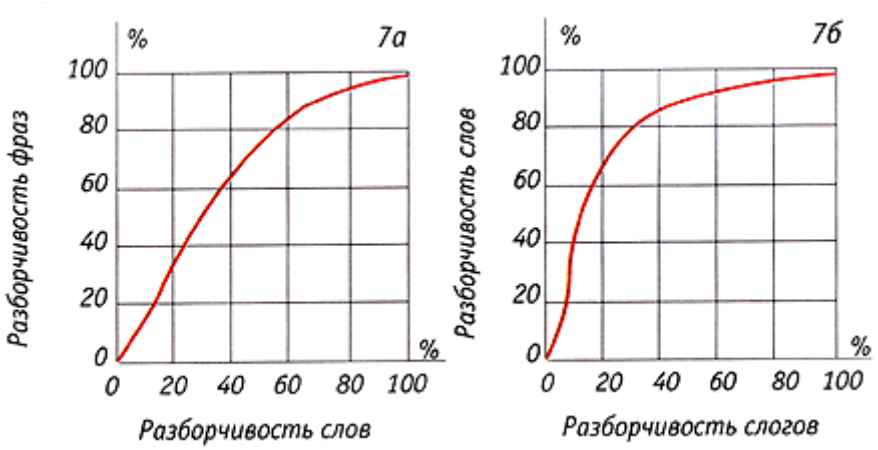

Рис. 7а - связь фразовой и словесной разборчивости;
7б - связь словесной и слоговой разборчивости

Большое влияние на результаты определения разборчивости речи оказывает не только количественный состав группы слушателей, но и другие факторы: образование, профессия, социальная принадлежность, а также память и сообразительность. Все слушатели должны обладать нормальным слухом, быть носителями данного языка, и должны быть знакомы со всеми тестовыми словами. Возрастной состав ограничен 35 годами. В процессе испытаний могут привлекаться как тренированная бригада экспертов, показания которой проверены на эталонной системе, так и нетренированные слушатели (при этом их количество должно быть больше).

Для ориентировочной оценки результатов испытаний в стандарте приведены классы средних значений разборчивости речи, указанные в таблице.

Класс	Условия	Средние значения слоговой разборчивости в %

I	отличные	Свыше 90
II	хорошие	от 80 до 90
III	удовлетворительные	от 70 до 80
IV	плохие	Ниже 70

Наряду с разборчивостью, часто указываются и другие субъективные факторы, влияющие на качество восприятия речи. К ним относятся: громкость речи, эхо, порхающее эхо, нарушение локализации, тембровые искажения, повышенный уровень шума и плохие акустические условия в зоне расположения источника звука. Следует заметить, что громкость, эхо и шум являются факторами, которые непосредственно определяют разборчивость речи и косвенно оцениваются при субъективной оценке разборчивости.

В отечественных стандартах по оценке качества передачи речи по каналам связи (ГОСТ Р 50840-95 и ГОСТ 51061-97) также используется измерение слоговой разборчивости речи методом артикуляционных испытаний, и измерение фразовой разборчивости при нормальном и ускоренном темпах произнесения. При этом отбор экспертов, выбор слоговых таблиц и методы статистической оценки происходят практически по тем же правилам, только количество экспертов составляет 4:5 человек. Требования к каналам связи высшего качества составляют не менее 93% слоговой разборчивости.

В международных стандартах, в частности ANSI S3.2-89, предлагается использовать пять дикторов и пять экспертов, удовлетворяющих указанным выше требованиям, но процедура предъявления речевого материала значительно сложнее.

Таким образом, процедура организации субъективных экспертиз по оценке разборчивости речи - дело сложное, длительное и достаточно дорогостоящее, хотя и наиболее достоверное. Поэтому за последние годы большое внимание было уделено созданию объективных методов оценки разборчивости, что позволило внедрить в практику целый ряд новых достаточно эффективных компьютерных методов расчета разборчивости речи в различных условиях.